PB-237 351

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM

MEASURING POLLUTANTS FOR WHICH NATIONAL AMBIENT AIR QUALITY STANDARDS HAVE BEEN PROMULGATED

Research Triangle Institute

PREPARED FOR
National Environmental Research Center

August 1973

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

EPA-R4-73-028-e
318034
August 1973

PB 237 351

Environmental Monitoring Series

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM

MEASURING POLLUTANTS FOR WHICH NATIONAL AMBIENT AIR QUALITY STANDARDS HAVE BEEN PROMULGATED
FINAL REPORT

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC 20460

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before compl

PB 237 351

1. REPORT NO.
EPA-R4-73-028-e

2.

4. TITLE AND SUBTITLE
Guidelines for Development of a Quality Assurance Program. Measuring Pollutants for Which National Ambient Air Quality Standards Have Been Promulgated. (Final Report)

5. REPORT DATE
August 1973

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Franklin Smith and A. Carl Nelson, Jr.

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Research Triangle Institute
Research Triangle Park, N. C. 27709

10. PROGRAM ELEMENT NO.
1HA327

11. CONTRACT/GRANT NO.
68-02-0598

12. SPONSORING AGENCY NAME AND ADDRESS
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D. C. 20460

13. TYPE OF REPORT AND PERIOD COVERED
Final

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Five method specific field usable guidelines for quality assurance for ambient air pollutants were produced under the contract and published separately. This document is not a field document. It supplies some background information used in the development of the five documents.

PRICES SUBJECT TO CHANGE

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Quality Assurance Quality Control Air Pollution		13H 14D 13B

18. DISTRIBUTION STATEMENT
Unlimited

19. SECURITY CLASS *(This Report)*
Unclassified

20. SECURITY CLASS *(This page)*
Unclassified

21. NO. OF PAGES
175

22. PRICE
6.25 - 2.25

EPA Form 2220-1 (9-73)

EPA-R4-73-028-e

GUIDELINES FOR DEVELOPMENT OF A QUALITY ASSURANCE PROGRAM

MEASURING POLLUTANTS FOR WHICH NATIONAL AMBIENT AIR QUALITY STANDARDS HAVE BEEN PROMULGATED

FINAL REPORT

by

Franklin Smith and A. Carl Nelson, Jr.

Research Triangle Institute
Research Triangle Park, North Carolina 27709

Contract No. 68-02-0598
ROAP No. 26BGC
Program Element No. 1HA327

EPA Project Officer: Dr. Joseph F. Walling

Quality Assurance and Environmental Monitoring Laboratory
National Environmental Research Center
Research Triangle Park, North Carolina 27711

Prepared for

OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

August 1973

ia

This report has been reviewed by the Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

TABLE OF CONTENTS

Section		Page
I	INTRODUCTION	1
	GENERAL	1
	PROGRAM OBJECTIVES	2
	PROGRAM APPROACH	2
II	MEASUREMENT METHODS	4
	GENERAL	4
	REFERENCE METHOD FOR THE DETERMINATION OF SUSPENDED PARTICULATES IN THE ATMOSPHERE (HIGH VOLUME METHOD)	6
	General	6
	Functional Analysis	7
	Arriving At Suggested Performance Standards	25
	Recommendations	30
	REFERENCE METHOD FOR DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)	32
	General	32
	Identification and Modeling of Important Parameters	33
	Performance Model	41
	Arriving at Suggested Performance Standards	44
	Recommendations	48
	REFERENCE METHOD FOR THE MEASUREMENT OF PHOTOCHEMICAL OXIDANTS CORRECTED FOR INTERFERENCES DUE TO NITROGEN OXIDES AND SULFUR DIOXIDE	49
	General	49
	Arriving at Suggested Performance Standards	49
	Recommendations	56

TABLE OF CONTENTS (Cont'd)

Section Page

TENTATIVE METHOD FOR THE CONTINUOUS MEASUREMENT OF NITROGEN DIOXIDE (CHEMILUMINESCENT) 57

General 57

Arriving at Suggested Performance Standards 57

Recommendations 58

REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY) 67

General 67

Arriving at Suggested Performance Standards 67

Recommendations 72

III BACKGROUND INFORMATION APPLICABLE TO THE MANAGEMENT MANUALS OF FIELD DOCUMENTS 73

GENERAL 73

SELECTING AUDITING SCHEMES 75

Computation of the Probability of Accepting a Lot as a Function of the Sample Size and the Acceptable Number of Defectives 75

Confidence Interval Estimate for the Percentage of Good Measurements in a Lot 77

Computation of the Percent Defective in the Reported Data 81

COST IMPLICATIONS 85

Bayesian Scheme for Computation of Costs 86

Computation of Average Cost 89

Cost Analysis 92

TABLE OF CONTENTS (Cont'd)

Section		Page
	Cost Trade-Off Procedures	97
	DATA QUALITY ASSESSMENT	103
	Data Presentation	103
	Assessment of Individual Measurements	104
	Overall Assessment of Data Quality	105
	Estimated Variance of Reported Data Using the Test Statistic: $s^2/\sigma^2 = \chi^2/f$	107
IV	RECOMMENDATIONS AND CONCLUSIONS	110
	REFERENCES	111
APPENDIX 1 -	COMPUTER SIMULATION	114
APPENDIX 2 -	SENSITIVITY ANALYSIS	128
APPENDIX 3 -	FREQUENCY DISTRIBUTIONS	139
APPENDIX 4 -	ESTIMATION OF THE MEAN AND VARIANCE OF VARIOUS COMPARISONS OF INTEREST	144
APPENDIX 5 -	CONTROL CHARTS	158

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Particulate Concentration and Flow Rate as Functions of Time	16
2	Symmetrical Diurnal Concentration Pattern	16
3	Sensitivity of Measured Concentrations to Various Parameters for Different Sampling Environments	24
4	Experimental Plan for Selected Factors	31
5	Estimated Standard Deviation of Analyzer Reading (volts) Versus NO_2 Concentration (ppm)	66
6	Probability of d Defectives in the Sample If the Lot (N = 100) Contains D% Defectives	78
7	Probability of d Defectives in the Sample If the Lot (N = 50) Contains D% Defectives	79
8	Percentage of Good Measurements Vs. Sample Size for No Defectives and Indicated Confidence Level	82
9	Percentage of Good Measurements Vs. Sample Size for 1 Defective Observed and Indicated Confidence Level Lot Size = 100	83
10	Average Cost Vs. Audit Level (Lot Size N = 100)	91
11	Added Cost ($) Vs. MSE (%) for Alternative Strategies	99
12	Precision of Reported Measurements Vs. Cost of Quality Control Procedure	100
13	Precision of Reported C vs. Total Added Cost, C_T	102
14	Critical Values of Ratio s_i/σ_i Vs. n	109
1-1	Simulated Distribution of Particle Concentration (Mean Concentration = 100 $\mu g/m^3$) and Normal Distribution Approximation	120
3-1	Sketch of Normal Frequency Distribution	140
3-2	Sketch of Uniform Frequency Distribution	143

LIST OF ILLUSTRATIONS (Concl'd)

Figure No. — Page

4-1 Comparison of Individual Analyzer Reading With Value From Calibration Curve 149

4-2 Confidence and Tolerance Limits for the Example 153

5-1 Control Chart for Means 159

5-2 Control Chart for Ranges 159

5-3 Subdivision of Control Chart Into Zones and Assigned Scores 161

5-4 Standard Control Chart Comparing Techniques 162

LIST OF TABLES

Table No. Page

1 PARAMETER EFFECTS MODELS AND VARIABLE MODELS AS USED IN SIMULATION AND SENSITIVITY ANALYSES 19

2 SYSTEM SIMULATION RESULTS 22

3 COLLABORATIVE TESTING DATA--HIGH VOLUME METHOD FOR PARTICULATES 26

4 SUGGESTED PERFORMANCE STANDARDS 28

5 PARAMETER EFFECTS MODELS AND VARIABLE MODELS AS USED IN SIMULATION AND SENSITIVITY ANALYSIS (SO_2) 40

6 SYSTEM SIMULATION RESULTS (SO_2) 42

7 SENSITIVITY ANALYSIS RESULTS (SO_2) 43

8 SUGGESTED PERFORMANCE STANDARDS 45

9 SUGGESTED PERFORMANCE STANDARDS 50

10 CALIBRATION DATA FOR CHEMILUMINESCENT OZONE MONITOR 51

11 ANALYSIS OF VARIANCE OF KI DATA 52

12 ANALYSIS OF VARIANCE OF CHEMILUMINESCENT MONITOR DATA 52

13 ANALYSIS OF VARIANCE OF KI DATA FOR EACH SLEEVE SETTING 54

14 ESTIMATE OF STANDARD DEVIATION FOR REPEATABILITY $(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)^{1/2}$ 54

15 ANALYSIS OF VARIANCE OF CHEMILUMINESCENT OZONE MONITOR DATA FOR EACH SLEEVE SETTING 55

16 ESTIMATE OF STANDARD DEVIATION FOR REPEATABILITY $(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)^{1/2}$ 55

17 SUGGESTED PERFORMANCE STANDARDS 59

18 DATA FOR ANALYZER A 60

19 DATA FOR ANALYZER B 61-62

20 ANALYSIS OF VARIANCE OF ANALYZER A DATA 63

21 ANALYSIS OF VARIANCE OF ANALYZER B DATA 64

LIST OF TABLES (Cont'd)

Table No. Page

22 SUMMARY OF DATA ANALYSES - ANALYZER A 65

23 SUMMARY OF DATA ANALYSES - ANALYZER B 65

24 SUGGESTED PERFORMANCE STANDARDS 68

25 P(d defectives) 77

26 COSTS VS. DATA QUALITY 86

27 COSTS IF 0 DEFECTIVES ARE OBSERVED AND THE LOT IS REJECTED 87

28 COSTS IF 0 DEFECTIVES ARE OBSERVED AND THE LOT IS ACCEPTED 87

29 COSTS IN DOLLARS 88

30 OVERALL AVERAGE COSTS FOR ONE ACCEPTANCE-REJECTION SCHEME 90

31 OVERALL AVERAGE COSTS 90

32 QUALITY OF INCOMING DATA AND ASSOCIATED AVERAGE COST 92

33 SUMMARY INFORMATION 93

34 EFFECT OF VARYING THE COST OF ACCEPTING POOR QUALITY DATA 94

35 EFFECT OF VARYING THE COST OF REJECTING GOOD QUALITY DATA 95

36 EFFECT OF VARYING THE PERCENT DEFECTS IN POOR QUALITY LOTS 95

37 ASSUMED STANDARD DEVIATIONS FOR ALTERNATIVE STRATEGIES 98

38 CRITICAL VALUES OF s_i/σ_i 108

1-1 INPUT DATA 116

1-2 PORTION OF SIMULATED DATA 117

1-3 CHECK OF SIMULATED VALUES 118

1-4 LISTING OF SIMULATED VALUES IN ASCENDING ORDER 119

1-5 MOMENTS OF SIMULATED DISTRIBUTION 121

LIST OF TABLES (Concl'd)

Table No.		Page
1-6	INPUT DATA (SO_2)	123
1-7	PORTION OF SIMULATED DATA (SO_2)	124
1-8	CHECK OF SIMULATED VALUES (SO_2)	125
1-9	LISTING OF SIMULATED VALUES IN ASCENDING ORDER (SO_2)	126
1-10	MOMENTS OF SIMULATED DISTRIBUTION (SO_2)	127
2-1	PARTIAL OUTPUT OF SENSITIVITY ANALYSIS	133
2-2	PRINTOUT OF SENSITIVITY ANALYSIS OF S.P.	134
2-3	PARTIAL OUTPUT OF SENSITIVITY ANALYSIS (SO_2)	137
2-4	PRINTOUT OF SENSITIVITY ANALYSIS (SO_2)	138
4-1	COMPUTATION OF CONFIDENCE AND TOLERANCE LIMITS FOR THE EXAMPLE	152
5-1	TABLE OF FACTORS FOR CONSTRUCTING CONTROL CHARTS	160

SECTION I INTRODUCTION

GENERAL

Surveillance of air quality is an important function of any air pollution control agency and an integral part of the total effort to control air pollution.

Air quality measurements are required to

(1) determine the presence and extent of air pollution problems within a given jurisdictional area,

(2) develop implementation plans,

(3) document progress towards attainment of standards and,

(4) provide immediate air quality data when air pollution episode conditions exist.

Air pollution control agencies at all levels of government, (i.e., federal state, and local) are currently monitoring a wide range of gaseous and particulate pollutants. Primary responsibilities for monitoring rest with state and local agencies. The Environmental Protection Agency (EPA) provides assistance through issuance of monitoring guidelines, development and field testing of instruments and analytical methods, promulgation of reference methods, issuance of recommended calibration, operation, and maintenance procedures for continuous analyzers, and training activities.

National primary and secondary ambient air quality standards, reference measurement methods, and sample averaging times have been promulgated for suspended particulates, sulfur dioxide, carbon monoxide, photochemical oxidants, and non-methane hydrocarbons. Regulations have been published requiring implementation plans which will allow achievement of these standards. These plans require the operation of air surveillance networks which routinely measure the specified pollutants.

In order to insure that the results reported by the various networks are valid and that non-compliance with standards will be detected when it occurs, the Quality Assurance and Environmental Monitoring Laboratory of the EPA has developed quality assurance programs and procedures applicable to each of the reference methods. Implementation of a quality assurance program will result in data that are more uniform in terms of precision and accuracy and will enable each monitoring network to continuously generate data that approach the highest level of quality possible for a particular measurement method.

PROGRAM OBJECTIVES

The objectives of this program were to provide guidelines for the development of quality assurance programs applicable to measuring pollutants for which National Ambient Air Quality Standards have been promulgated. Specifically, guidelines were written for the reference methods for measuring suspended particulates, carbon monoxide, photochemical oxidants, sulfur dioxide, and one tentative method for the continuous measurement of nitrogen dioxide (chemiluminescent).

The objectives of a quality assurance program for any measurement method are to:

(1) provide routine indications of unsatisfactory performance of personnel and/or equipment,

(2) provide for prompt detection and correction of conditions which contribute to the collection of poor quality data, and

(3) collect and supply information necessary to describe the quality of the data.

To accomplish the above objectives, the guidelines for each quality assurance program include directions for the:

(1) routine monitoring of the variables and parameters which may have a significant effect on data quality,

(2) development of statements and evidence to qualify data and detect defects,

(3) evaluation of relevant action strategies to vary the level of precision/accuracy in the reported data as a function of cost, and

(4) routine training and evaluation of operators.

PROGRAM APPROACH

The program approach is discussed in terms of the content, the means of prescribing the guidelines to achieve maximum utility, and scope of the guidelines for each method.

In presenting quality assurance guidelines, an individual, field-useable document was written for each method. Backup data, analyses, and discussions to further elucidate, in a more rigorous manner, the contents of the field documents are included in this final report.

Each field document consists of three major parts. They are:

(1) OPERATIONS MANUAL - The Operations Manual sets forth recommended operating procedures, instructions for performing control checks designed to give an indication or warning that invalid or poor quality data are being collected, and instructions for performing certain special checks for auditing purposes.

(2) SUPERVISION MANUAL - The Supervision Manual contains directions for 1) assessing air quality data, 2) collecting information to detect and/or identify trouble, 3) applying quality control procedures to improve data quality, and 4) varying the auditing or checking level to achieve a desired level of confidence in the validity of the outgoing data. Also, monitoring strategies and costs as discussed in the Management Manual are summarized in this manual.

(3) MANAGEMENT MANUAL - The Management Manual presents procedures designed to assist the manager in 1) detecting when data quality is inadequate, 2) assessing overall data quality, 3) determining the extent of independent auditing to be performed, 4) relating costs of data quality assurance procedures to a measure of data quality, and 5) selecting from the options available the alternative(s) which will enable one to meet the data quality goals by the most cost-effective means. Also, discussions on data presentation and personnel requirements are included in this manual.

This final report contains sections on:

(1) the individual measurement methods which include any data and analyses used to arrive at suggested performance standards in the respective field documents and discussions of the methods used to treat areas where sufficient field data were not available.

(2) background information directly applicable to the Management Manual sections in the field documents in the form of a broader and more rigorous mathematical treatment of the subject areas in that manual, and

(3) recommendations for implementation of quality assurance programs, areas requiring further study, and conclusions arrived at as the result of this program.

SECTION II MEASUREMENT METHODS

GENERAL

A systems approach was employed in writing guidelines for developing a quality assurance program for each measurement method. The approach was basically the following.

A functional analysis of the method was performed to identify and, if possible, quantify sources of variability. Total variability of a measurement process is a combination of two error components; namely, determinate and indeterminate. Determinate errors are defined as those that can be avoided once they are recognized. This type of error is caused by such factors as:

(1) improper calibration of glassware or instruments, or improper standardization of reagents;

(2) personnel errors, such as the tendency of an operator to read a meter too high or too low; and

(3) a constant error in the method.

Determinate errors introduce a bias into the measured values. Indeterminate errors, or chance variations, have a random nature and cannot be completely eliminated from a measurement process, but can possibly be contained in a narrow zone when their sources are well known.

The second step in writing the guidelines was to delineate operating procedures, quality control checks, and suggested performance standards designed to eliminate determinate errors and to minimize indeterminate errors. Also, independent audit checks encompassing all phases of the measurement process were presented to allow for an independent estimate of data quality in terms of precision and accuracy.

The third and final step was to evaluate certain monitoring strategies for improving data quality and different auditing schemes to vary the level of confidence in the estimated quality of the reported data as functions of cost.

In all cases suggested performance standards and control limits were arrived at from results of a collaborative test of the particular method, published data from special tests, or from engineering judgment. All such values must be evaluated and adjusted as field data become available. To employ any said value as a "hard" standard without first evaluating its validity for field conditions would be contrary to the intent of the authors and would, in their opinion, significantly reduce the utility of the documents.

Suggested performance standards given in the individual field documents, which were derived from actual data, are treated in detail in this report. Suggested standards which resulted from engineering judgments or estimates are adequately described in the field documents and are not addressed in this report. Suggested standards for auditing and for operation are somewhat common to all methods and are summarized here and not repeated for each method.

Audit rates are discussed in the field documents and in Section III of this report. For normal, routine monitoring purposes an auditing level of 7 checks out of a lot size of 100 samples was recommended. For the conditions spelled out in the field document this level of auditing allows one to say at the 50 percent confidence level that:

(1) If no defects are found in the 7 checks, at least 90 percent of the samples in the lot are good (read from the 50% curve in Figure 8, page 82, for a sample size of 7).

(2) If no more than 1 defect is observed in the 7 checks at least 76 percent of the samples in the lot are good (read from the 50% curve in Figure 9, page 83, for a sample size of 7).

This turned out to be the most cost effective auditing level for the costs assumed in Section III, under Computation of Average Cost, page 89.

Suggested standards for operation involve evaluating auditing data to determine if the measurement process is "in" or "out" of control. In all cases the suggested standards for defining defects were given at the 3σ level in the field documents. It was assumed that the data--that is, the difference between the value derived from the audit and the operator's measured value--are normally distributed and that the means and standard deviations as estimated or computed from a limited amount of test data represent the true mean and standard deviation of the total population in question (see Appendix 3 for a detailed discussion of the normal distribution). Under these assumptions and for an auditing level of 7 checks out of a lot size of 100, the measuring process was judged to be "out" of control and in need of trouble shooting and corrective actions if any of the following conditions were observed:

(1) One check value exceeds the ± 3σ limits (99.7 percent of the checks should be within ± 3σ of the mean; in this case .997 × 7 = 6.979, or approximately all 7 checks, should be within ± 3σ if the process in "in" control).

(2) Two checks exceed the ± 2σ limits (under normal conditions 95 percent of the checks would be within ± 2σ of the mean).

(3) Four checks out of 7 exceed the ± 1σ limits (under normal conditions 68 percent of the checks, or 4.76, checks should be within ± 1σ of the mean).

Each measurement method is discussed separately in the following subsections.

REFERENCE METHOD FOR THE DETERMINATION OF SUSPENDED PARTICULATES IN THE ATMOSPHERE (HIGH VOLUME METHOD)

General

Measurement of the mass of suspended particulate matter in the ambient atmosphere by the High Volume Method requires a sequence of operations and events that yields as an end result a number that serves to represent the average mass of suspended particulates per unit volume of air over the sampling period. Techniques for dynamic calibration of high volume samplers using test atmospheres containing known concentrations of particulates are not available. Therefore, there is no way of knowing the accuracy of the values derived from high volume sampling. However, numerous experiments and studies have been performed to identify and evaluate factors which influence the final results. Also, a collaborative study has been made to determine the repeatability and reproducibility of the method under controlled conditions (Ref. 1).

The High Volume Method is dependent on several parameters such as operator effects, environmental conditions, calibration procedures, variation in instrumentation, and other variables and effects, some of which may be unknown at this time. With the relatively large number of parameters/variables involved, it was felt necessary to develop a mathematical model (hereafter referred to as a performance model) of the process which could be subjected to statistical analysis in order to:

(1) generate accuracy data for comparison with published data as a check on the reasonableness of essential assumptions made concerning important variables,

(2) estimate the process variability for various monitoring strategies,

(3) rank the variables according to their influence on the measured concentration, and

(4) evaluate the effect of various control procedures on the measurement process.

The development and use of the performance model are discussed in the following paragraphs as part of a functional analysis of the measurement process.

A functional analysis of the High Volume measurement process starting with the specification of the filter media and extending through the documentation and data reporting steps was performed to identify all parameters, variables, and errors which could be possible contributors to the variation in the measurements. Data sufficient to allow an estimate of the error range associated with each parameter in the measuring process were collected from published documents, private communications with EPA personnel, and from RTI in-house projects. If data were not available, engineering judgements were used to estimate variable limits. These data were used in developing a performance model for computer simulation.

The performance model was designed to account for the major parameters in the application of the method and to characterize their variations. It was subjected to statistical analysis by means of a computer program designed to introduce the kinds of variations to be expected of each of the parameters and to execute simulation for a sequence of different choices of conditions so that the effects of each parameter and any significant interactions between the parameters would be identified and evaluated. Results from the model study provided a basis for decisions concerning the quality control technology applied to the different steps in the measurement procedure and the suggested performance standards given in the field document.

The functional analysis is discussed in terms of parameter evaluation and modeling, the performance model, and results from the simulation and sensitivity analyses. Discussions of the computer programs written to perform the simulation and sensitivity analyses are given in Appendices 1 and 2, respectively.

Functional Analysis

A functional analysis consisting of identification and modeling of important variables and parameters, development of a performance model of the measurement process, and simulation and sensitivity analyses using the performance model was performed as a means of evaluating the overall measurement process. Each phase of the analysis is discussed below.

Identification and Modeling of Important Parameters - The parameters are grouped according to whether they influence particulate weight, flow rate, sampling time, or the measured concentration directly.

Factors which could influence particulate weight include: 1) filter surface alkalinity, 2) relative humidity of the filter conditioning environment, 3) elapsed time between sample collection and analysis, and 4) weighing errors.

Factors influencing the precision and accuracy of the calculated average flow rate are: 1) method of calculating the average flow rate, 2) flow rate calibration, and 3) temperature and pressure effects on flow rate.

Sampling time is usually measured by an elapsed time indicator operated from an electronic timer capable of being pre-set to start and stop the sampler.

One factor affects the measured concentration directly; that is, unequal sampling rates. Unequal sampling rates result when particulate concentration and sample air flow vary with time.

Filter surface alkalinity - Flash fired glass-fiber filters are the most frequently used filters for collecting suspended particulate matter for gravimetric analysis. It has been shown (Refs. 2-4) that solid matter is deposited on the fiber surfaces by oxidation of acid gases in the sample air. It was also observed that the quantity of such matter deposited in a given sampling period was not the same for all commercially available glass-fiber filters. Although other reactions are conceivable, the formation of sulfate was studied. It occurs during the first 4 to 6 hours of sampling, and very little is formed after 6 hours (Ref. 2).

Tests conducted with filters of pH-6.5 and pH-11 showed a significantly larger sulfate to total particulates ratio for the filters of pH-11 (Ref. 3). Additional tests (Ref. 4) have shown that alkaline filter media can yield erroneously high results for total particulate matter, sulfates, nitrates, and other species existing as acid gases in the sample air. Samplers operating side by side, one equipped with a filter of pH-11 and the other with a filter of pH-6.5, showed after 9 sampling periods that the average total particulate matter was higher by 18 percent, sulfates by 40 percent, and nitrates by 60 percent for filters of pH-11 than for filters of pH-6.5.

The quantity of solid matter deposited during a sampling period is a function of filter pH, length of sampling period or volume of air sampled, and the concentration of acid gases in the sample air. However, even background levels of NO_2 and SO_2 well below national air quality standards can induce significant errors when alkaline filters are used.

In modeling this variable, the following points were considered. First, the effect can be controlled or eliminated by specifying neutral filters when ordering a batch of filters. Also, the data from references 3 and 4 show a large difference in the magnitude of the effect. Using data from either reference independently would result in the effect being insignificantly small (Ref. 2) or being so large as to mask the effects of other parameters (Ref. 4). The values used to model the pH effect, E(pH), for this analysis were selected so that the effect will always show up as being significant but still not mask the effects of other parameters. Also, it was assumed that there are always sufficient concentrations of acid gases in the atmosphere to effect this reaction.

The pH effect was modeled as a function of filter pH and of sampling period (T) time in minutes by:

$$E(pH) = 33.5\ (pH-7)(T)^{1/2}$$

where E(pH) = weight of acid gases converted to particulate matter during the sampling period in μg,

pH = the pH of the filter, and

T = sampling time in minutes.

This model shows a zero effect for a neutral filter (pH-7) and a conversion of 5,000 μg (5 mg) for a filter of pH-11 and a 24-hour (1440-minute) sampling period.

For simulation purposes in guideline development, values of pH were generated from a uniform distribution ranging from 7 to 12 (see Appendix 3 for a discussion of the uniform distribution). However, in an actual application the pH value or range should be determined and used in the model.

Relative humidity effect - Collected particulates are hygroscopic in varying degrees. Samples collected from suburban, urban, and industrial atmospheres were weighed after being conditioned for a minimum of 4 days at relative humidities varying from 0 to 100 percent (Ref. 5). The results show less than a 1 percent increase in particulate weight in going from 0 to 55 percent relative humidity. However, the relationship is exponential for relative humidities greater than 55 percent, showing a 5 percent increase in particulate weight at a relative humidity of 70 percent and approximately 15 percent weight increase at 80 percent relative humidity. The industrial sample proved most hygroscopic with a 90 percent weight increase at a relative humidity of 100 percent.

The above results point out the importance of maintaining the conditioning environment at a relative humidity less than 55 percent. Also, the humidity level should be the same for conditioning the exposed filter as that used to condition the clean filter. In instances in which the exposed filter has to be removed from the conditioning environment for weighing, the time interval between removal and weighing should be kept to a minimum. An interval of less than 5 minutes is recommended.

It must be remembered that the exposure period in this test was 4 days. Also, in filter-conditioning environments the relative humdity is probably never over 60 to 65 percent. It also seems unlikely that samples would be out of the conditioning environment for over an hour, if at all, before weighing. Therefore, it appears reasonable to assume that under normal operating conditions the humidity effect would seldom exceed that of a 24-hour exposure to a 60 to 65 percent relative humidity. The relative humidity effect was modeled as

$$E(RH) = (W_p)\ 0.00013 \exp(0.087 \times RH)$$

where $E(RH)$ = increased particulate weight due to adsorption of moisture in μg,

W_p = true weight of collected particulates in μg, and

RH = relative humidity of the conditioning environment expressed as a percent.

For simulation purposes the relative humidity was modeled as a normal distribution with a mean of 50 and a standard deviation of 5 (see Appendix 3 for discussion of the normal distribution). This would result in an error of 1 percent if the exposed filter was conditioned at 50 percent RH and a 2.7 percent error if conditioned at 60 percent RH.

Elapsed time between sample collection and analysis - During the time between sample collection and final weighing volatile matter having substantial vapor pressures may evaporate resulting in a significant reduction in particulate weight.

Results from one set of tests (Ref. 6) indicate that the weight loss is approximately proportional to the percent of organic matter initially present in the collected sample. The greatest rate of loss is experienced during the first 24 hours after collection. A lower but somewhat constant rate of loss continues for several days, the number of which is again a function of the initial content of organic matter.

The test involved samples collected from three different sources. Particulates from these sources averaged 1.8, 9.3, and 61.5 percent initial organics. Weight losses were measured at elapsed times of from less than 24 hours up to 26 days for several samples.

In modeling the results of this test, the weight of the particulates after the 24-hour conditioning period was taken as the reference weight. That is, no error was associated with the losses occuring during the first 24 hours after sample collection. This approach was taken because a 24-hour conditioning period is specified by the reference method and the manner in which the above test was conducted may have biased the indicated weight losses during this initial 24-hour period. For example, the samples were placed in a desiccator for only 30 minutes to remove extraneous moisture. The sample was stabilized for another 30 minutes in a controlled environment at 75° ± 3°F and 54 ± 3 percent relative humidity, and then weighed. Thus, the conditioning period was only 1 hour. The high rate of weight loss during the first 24 hours may have been due to increased loss of moisture as well as the evaporation of organics.

Weight loss due to evaporation of organic matter was modeled as

$$E(OM) = W_p \, [0.000054 \, (\% \text{ of OM})(D)]$$

where $E(OM)$ = weight loss in μg,

W_p = true weight of collected particulates in μg,

% of OM = initial organic content of collected particulates in percent, and

D = time between collection and weighing in days minus 1 day.

This model gives a weight loss of 1 percent per 10 percent organic content for a delay of 12 days and a zero loss when no delay is involved.

For simulation purposes the initial organic content was modeled as a normal distribution with a mean of 20 and a standard deviation of 5 (see Appendix 3 for definitions of normal distribution). Values for delays in days were randomly selected from a uniform distribution with a range of 0 to 21.

Weighing errors - Two weighing processes are involved in the High Volume Method. They are the weighing of clean filters and the weighing of exposed filters. If not properly monitored, the weighing process can be a source of significant error in the final result derived from the High Volume Method. Fifty tare-weight weighings for each of five filters made over a period of time in which the relative humidity of the conditioning chamber was varied from 20 to 50 percent showed a maximum variation in tare-weight weighings of 1.2 mg (Ref. 5). Another test showed a standard deviation of approximately 0.8 mg for weighing clean filters after successive 24-hour conditioning periods.

These data point out the importance of performing the weighings at the appropriate time, i.e., just after the 24-hour conditioning period, and the necessity of performing the audit or check within a few minutes either before or after the regular weighing in order to expect good agreement between the two weighings.

It is suggested that if the weighing and auditing procedures are properly carried out, the variation between the original and check weights of clean filters should not exceed ± 1.0 mg and not more than ± 2.7 mg for exposed filters. The combined weighing error is modeled as a normal distribution with a zero mean and a standard deviation of 1000 μg symbolized by

$$E(W) = N(0,1000)$$

where $E(W)$ = weighing error in μg, and

$N(0,1000)$ = represents a normal distribution with a zero mean and a standard deviation of 1000 μg (the accepted notation for a normal distribution is $N(\mu,\sigma)$.

Flow-rate reading error - The general feeling of people reading the rotameter is that they can read it to within ± 0.03 m^3/min (± 1 ft^3/min). For an average flow rate of 1.13 m^3/min (40 ft^3/min), this would represent a ± 2.5 percent error. Under field conditions the largest component of reading error is probably due to operator technique (e.g., parallex error, rotameter not held in a true vertical position, or estimating an average reading while the float is in oscillatory motion) and, if not monitored, would in some instances be much larger than 2.5 percent of the true value.

Flow-rate reading error was modeled as

$$E(Q_r) = N(0,0.025)$$

where $E(Q_r)$ = flow-rate reading error as a decimal fraction, and

$N(0,0.025)$ = a normal distribution (see Appendix 3) with a zero mean and a standard deviation of 0.025.

Calculating average flow rate - Calculating the average flow rate from initial and final values assuming a constant rate of change throughout the sampling period can result in large errors. One report (Ref. 7) shows an average bias ranging from -1.2 to +8.1 percent of the true average flow rate for 3 sets of data with 6 samples each. Bias is defined as the difference in average flow rate as computed from the initial and final measurements compared to the average derived from several measurements made throughout the sampling period. These errors can result from particulates plugging the filter resulting in a nonuniform decrease in the flow rate over the sampling period or from variations in source voltage. . Nonuniform changes in flow rate are probably greatest in industrial areas due to sticky particulates and can result in a -2 to +10 percent error range in average flow-rate values.

A sampler equipped with a continuous flow-rate recorder does not have the above problem. The true average flow rate can be estimated or calculated by hourly values to within 0.03 m^3/min (1 ft^3/min) from the recorder chart. This represents a significant improvement in system accuracy.

For simulation purposes the error in computing the average flow rate using initial and final flow-rate measurements and assuming a constant rate of change was modeled as a function of the change in flow rate by the relationship

$$E(\bar{Q}) = \frac{\Delta Q}{0.5\ Q_i}\ [U(-0.02 \text{ to } +0.10)]$$

where $E(\bar{Q})$ = the error in Q in percent,

ΔQ = change in flow rate for the sampling period in m^3/min/24-hour

Q_i = initial flow rate in m^3/min, and

$U(-0.02 \text{ to } +0.10)$ = uniform distribution with lower limit of -0.02 and upper limit of +0.10. The notation for uniform distributions used in this document is U (lower limit to upper limit), see Appendix 3 for a discussion of the uniform distribution.

The model is such that samples experiencing a change in flow rate as great as 50 percent of the initial flow rate would be subject to biases ranging from -2 percent to +10 percent of the average flow rate. As the magnitude of change in flow rate decreases, the magnitude of the possible error decreases also. It should be pointed out that the bias may be negative. In actual practice the final flow rate may be greater than the initial flow rate due to voltage variations or extreme changes in the relative humidity during the sampling period.

The final flow rate, Q_f, was allowed to vary according to a uniform distribution (see Appendix 3 for discussion of uniform distribution) with a range of from 1 to 1.7 m^3/min resulting in a ΔQ range of from 0 to 0.7 m^3/min/24-hour. For a specific situation the drop in flow rate should be directly related to the total weight of particulates collected. However, for a large population of sites it was felt that the type of particulates were more important. That is, large flow-rate drops can result from a relatively small quantity of sticky particulates.

Flow-rate calibration - Calibration of 12 new orifice units by well-qualified individuals using positive displacement meters as primary standards under laboratory conditions showed a standard deviation from the mean of 2.1 percent (Ref. 1). Calibration of samplers in the laboratory and in the field by less qualified people using the orifice unit would be expected to yield a much larger standard deviation. Previous experience with high volume samplers indicates that +3 percent of the mean is a reasonable value to use as a standard deviation for calibration error for a well-monitored operation. There is also a possible change in the calibration with time. Such a term is not incorporated in this treatment but could be once sufficient field data are available to allow an estimate of its magnitude and characteristics.

For simulation purposes the calibration effect was assumed to be normally distributed with a zero mean and a standard deviation of 3 percent of the flow rate, i.e.,

$$E(Q_c) = N(0,0.03)$$

where $E(Q_c)$ = calibration error expressed as a decimal fraction, and

$N(0,0.03)$ = a normal distribution (see Appendix 3) with a zero mean and standard deviation of 0.03.

Temperature and pressure effects on flow rate - In most regions in the United States and for a specific site location, temperatures usually range from -4°C (25°F) to 38°C (100°F) and barometric pressure variations are on the order of ± 12.7 mmHg (0.5 in. Hg) (Ref. 8). Tests on a sampler equipped with a flow-rate recorder showed a maximum deviation from the calibration curve of + 7 percent to -10 percent in the indicated flow rate when going from the extremes of 38°C and 736 mmHg to -4°C and 762 mmHg. Calibration conditions were 70°F and 750 mmHg.

The above data point out the need for either calibrating on site or making corrections for temperature and pressure if the ambient site conditions are significantly different from the laboratory conditions.

No similar data were available for a sampler equipped with a rotameter, but it was felt that the variability would be at least as large. The temperature and pressure effect was modeled as a normal distribution with a zero mean and standard deviation of 3 percent of the measured flow rate, i.e.,

$$E(Q_{BT}) = N(0,0.03)$$

where $E(Q_{BT})$ = error in measured flow rate due to changes in temperature and pressure as a decimal fraction, and

$N(0,0.03)$ = normal distribution (see Appendix 3) with a zero and standard deviation of 0.03.

The error levels from this model would be less than ± 9 percent 99.7 percent of the time, less than ± 6 percent 95 percent of the time, and less than ± 3 percent 68 percent of the time.

This model is only valid for a specific site where the sampler has been calibrated at or near the average temperature and pressure. It does not represent the error that would result if the sampler were calibrated at a given temperature and pressure then used in a location where the average temperature or pressure is significantly different from the calibration conditions. Such a situation would require a bias term--that is, the mean of the distribution would no longer be zero.

Sampling time - The results of high volume sampling are not very sensitive to the normal magnitudes of timing errors. For example, a 14-minute error in a 24-hour sampling period results in a 1 percent error in the measured concentration. The reference method specifies that start and end times be determined to the nearest 2 minutes. This can be accomplished with the operators' watch or by using an elapsed time indicator on the sampler. In the first instance there is no way of knowing of or compensating for power failures or other interruptions occurring during the sampling period.

This could result in significant error in the sampling period time even though the start and end times are accurate. Samplers equipped with an elapsed time indicator or a continuous flow-rate recorder would indicate such power interruptions and allow for corrections to be made.

In generating simulation and sensitivity data, the error in sampling period time was assumed to be normally distributed about a mean value of zero with a standard deviation of 7 minutes. The model is

$$E(T) = N(0,7)$$

where $E(T)$ = error in sampling period time in minutes, and

$N(0,7)$ = normal distribution (see Appendix 3) with a zero mean and a standard deviation of 7 minutes.

Unequal sampling rates - In certain instances, when both flow rate and particulate concentration vary during the sampling period, significant errors in the measured average concentration can occur. The example given in Figure 1 is proposed as an extreme condition which is not likely to be exceeded in actual operations. In this case the concentration of suspended particulates varies from 353 $\mu g/m^3$ (10 $\mu g/ft^3$) to 70.6 $\mu g/m^3$ (2 $\mu g/ft^3$) according to the following equation

$$S.P. = 141.2\left(\frac{3}{2} + \cos\frac{\pi}{24}t\right)$$

where S.P. is the instantaneous concentration in $\mu g/m^3$, and

t is the time in hours.

Also, the flow rate decreases from 1.7 m^3/min (60 ft^3/min) to 1.02 m^3/min (36 ft^3/min) in a linear fashion according to the following relationship

$$Q = 1.7 - (0.03\ m^3/hr)t$$

where Q is the flow rate in m^3/min, and

t is the time in hours.

The true average concentration, $\overline{S.P.}$, is seen to be the value at the point where the concentration curve crosses 12 on the time axis, or

$$\overline{S.P.} = \frac{\int_{t_1}^{t_2} S.P.\ dt}{t_2 - t_1} = \frac{\int_{0}^{24} 141.2\left(\frac{3}{2} + \cos\frac{\pi}{24}t\right)dt}{24} = 212\ \mu g/m^3\ (6\ \mu g/ft^3).$$

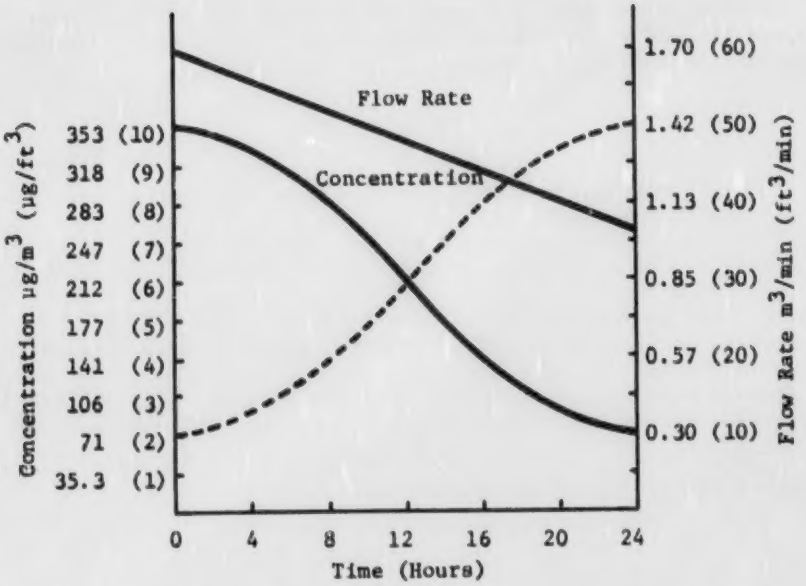


Figure 1: Particulate Concentration and Flow Rate as Functions of Time

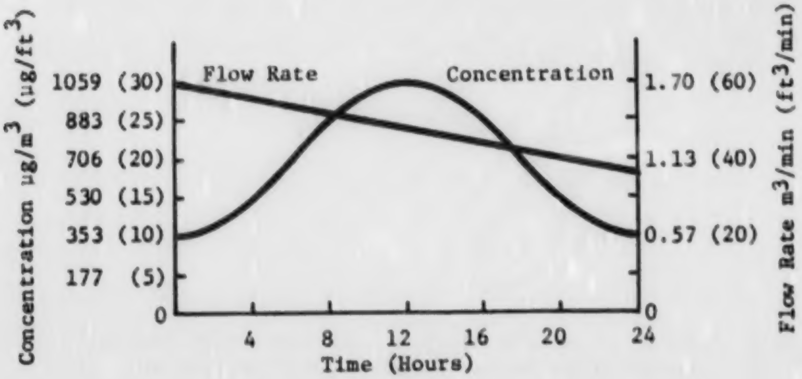


Figure 2: Symmetrical Diurnal Concentration Pattern

However, since the flow rate also varies with time, the average concentration that would be measured by the high volume sampler, assuming no other errors are involved, is expressed as:

$$\overline{\text{S.P.}}' = \frac{\int_{t_1}^{t_2} \text{S.P. } Q dt}{\int_{t}^{t_2} Q\, dt} = \frac{\int_{o}^{24} 141.2\left(\frac{3}{2} + \cos \frac{\pi}{24}\, t\right)(1.7 - 0.03\, t)dt}{\int_{o}^{24} (1.7 - 0.03\, t)dt}$$

$$= 227\ \mu g/m^3\ (6.42\ \mu g/ft^3).$$

This value differs from the true average concentration by + 7 percent. The reverse case in which the concentration increases as the flow rate decreases is illustrated by the dashed curve in Figure 1 and results in a -6.7 percent deviation from the true average.

In a situation such as that shown in Figure 2 in which the concentration exhibits a diurnal pattern which is symmetrical about the midpoint of the sampling period, the true average concentration is realized by the High Volume Method as long as the flow rate is a linear function of time. In this case

$$\text{S.P.} = 353\ [2 + \sin\left(\frac{\pi}{12}\, t + \frac{3}{2}\right)]$$

and

$$Q = 1.7 - (0.03\ m^3/hr)t\ .$$

Performing the same calculations as those done in the example in Figure 1 shows that the "measured" value is the same as the true value.

As a result of the above calculations, it is felt that a deviation greater than ± 7 percent from the true average concentration due to this effect alone should be very rare. An estimate of the possible error for a given site could be made by using the local diurnal pattern of suspended particulate concentration and normal or average drop in the flow rate over a 24-hour sampling period to perform the above calculations. The error would not be significant unless the change in flow rate is greater than 20 percent of the initial flow rate, the diurnal pattern is extremely nonsymmetrical about the midpoint of the sampling period, and the maximum concentration is at least four times as great as the minimum.

For simulation and sensitivity analyses the effect due to unequal sampling rates was modeled as

$$E[Q(t),\ S.P.(t)] = S.P.\ \frac{\Delta Q}{0.5\ Q_i}\ [U(-0.09 \text{ to } +0.09)]$$

where $E[Q(t),\ S.P.(t)]$ = concentration error due to unequal sampling rates in $\mu g/m^3$,

S.P. = true concentration in $\mu g/m^3$,

ΔQ = change in flow rate in $m^3/min/$[illegible]-hour,

Q_i = initial flow rate in m^3/min, and

$U(-0.09 \text{to} +0.09)$ = uniform distribution over the range -0.09 to +0.09 (see Appendix 3 for discussion of the uniform distribution).

This is consistent with the calculated results, namely, that a maximum error of ± 7 percent in the average concentration may result when the change in flow rate is as great as $0.4\ Q_i$, depending on the variability in the particulate concentration. It also shows that the magnitude of the possible error decreases as the change in flow rate becomes smaller.

Summary of parametric and variable modeling - Parameters and variable relationships and models as used in the simulation and sensitivity analyses are summarized in Table 1.

Performance Model - A mathematical performance model of the measurement process incorporating the error terms discussed in the previous subsection and summarized in Table 1 was derived from the basic equation

$$S.P. = \frac{W}{V}$$

where S.P. = average concentration in $\mu g/m^3$,

W = measured weight of collected particulates in μg,

and V = the measured volume of air sampled in m^3.

Table 1: PARAMETER EFFECTS MODELS AND VARIABLE MODELS AS USED IN SIMULATION AND SENSITIVITY ANALYSES

Parameter Effects	Variable Models
$E(pH) = 33.5\ (X(1) - 7)\ (T)^{1/2}$	$X(1) = pH = U(7.0 \text{ to } 12.5)$
$E(RH) = N^{*}(45, 5)$	$X(2) = E(RH)$
$E(W) = N(0, 1000)$	$X(3) = E(W)$
$E(Q_r) = N(0, 0.025)$	$X(4) = E(Q_r)$
$E(\bar{Q}) = \frac{Q_i - X(13)}{0.5\ Q_i}\ [X(5)]$	$X(5) = U^{**}(-0.02 \text{ to } +0.10)$
	$X(13) = Q_f = U(1 \text{ to } 1.7)$
$E(Q_c) = N(0, 0.03)$	$X(6) = E(Q_c)$
$E(Q_{BT}) = N(0, 0.03)$	$X(7) = E(Q_{BT})$
	X(8) extra variable, set equal to zero for computer runs
$E[Q(t),C(t)] = S.P.\ \frac{Q_i - X(13)}{0.5\ Q_i}\ [X(9)]$	$X(9) = U(-0.09 \text{ to } +0.09)$
$E(OM) = W_p[0.000054\ (X(10))(X(11))]$	$X(10) = D = U(1 \text{ to } 21)$
	$X(11) = \%\ OM = N(20, 5)$
$E(T) = N(0, 7)$	$X(12) = E(T)$

[*] $N(\mu,\sigma)$ represents a normal distribution with a mean μ and standard deviation σ, see Appendix 3.

[**] U(a to b) represents a uniform distribution with lower limit a and upper limit b, see Appendix 3.

For an actual measurement where error terms are involved, the measured particulate weight, W, would be

$$W = 0.5\ (Q_i + Q_f)\ (S.P.)_T + E(W) + E(pH) + E(RH) - E(OM)$$

where $0.5\ (Q_i + Q_f)\ (S.P.)_T$ = the weight of particulates that would be collected for a true concentration, $(S.P.)_T$, of suspended particulates if no errors were involved.

The remaining error terms are as described in the previous subsection.

The measured volume is

$$V = \bar{Q}\ T_m$$

where $\bar{Q}$ = the measured average flow rate in m^3/min given by

$$\bar{Q} = 0.5(Q_i + Q_f)\ [1 + E(Q_r) + E(\bar{Q}) + E(Q_c) + E(Q_{BT})]$$

and T_m = the measured sampling period time in minutes given by

$$T_m = T_t + E(T).$$

The term $0.5(Q_i + Q_f)$ represents the true average flow rate, and T_t is the true sampling period time.

Combining the above relationships plus the error due to unequal sampling rates, i.e., E[Q(t), S.P.(t)], gives the following

$$S.P. = \frac{0.5(Q_i+Q_f)\ (S.P.)_T + E(W) + E(pH) + E(RH) - E(OM)}{0.5(Q_i+Q_f)[1 + E(Q_r) + E(\bar{Q}) + E(Q_c) + E(Q_{BT})][T_t+E(T)]}$$

$$+ E[Q(t),\ S.P.(t)]$$

where S.P. is the "measured" average concentration in $\mu g/m^3$ for a sampling period.

The performance model was programmed for the Bunker Ramo 340 computer and simulation and sensitivity analyses performed as described in Appendices 1 and 2, respectively.

Simulation Results - Simulation results for 6 computer runs are summarized in Table 2. Computer runs 1 and 2 with input concentrations of 200 and 100 $\mu g/m^3$, respectively, are representative of the results obtainable from a large population of samples with parameters and variables modeled as in Table 1. The results show no overall bias, i.e., the average measured value is the same as the true value, and standard deviations of 7.5 and 7.2 percent of the mean for mean values of 200 and 100 $\mu g/m^3$, respectively.

Computer runs 3 and 4 were made with the major bias terms controlled. That is, neutral filter media is assumed eliminating the E(pH) term and a pressure transducer and continuous recorder replaces the rotameter reducing the possible error in estimating the average flow rate, $E(\bar{Q})$, and flow-rate reading error, $E(Q_r)$. The results show a zero bias and a standard deviation of 5 percent of the mean value for both concentration levels. This is a reduction of 16 percent or greater in the standard deviations of runs 1 and 2.

Computer run 5 indicates what could be achieved if all quality control checks recommended in the field document were followed, i.e., a continuous flow-rate recorder is used, and there is no delay between sample collection and analysis other than the 24-hour conditioning period. This approximates what is felt to be the best data quality achievable under field conditions. The results show a standard deviation of 4 percent of the mean value and a slight positive bias (1 percent).

Simulation results do not substantiate the accuracy value given for the reference method (Subsection 4.2, page 8191, of the Federal Register, Vol. 36, No. 84, Part II, Friday, April 30, 1971). The stated value of ± 50 percent of the true average concentration appears to be based primarily on the results of a computation of error due to unequal sampling rates performed in Reference 9. That computation is in error. Unequal sampling rates alone should seldom result in inaccuracies greater than ± 7 percent of the true average according to the calculations in the subsection on "Unequal sampling rates" starting on page 15 of this document. Simulation results indicate that a standard deviation of about 7 percent, or less, of the true value should be easy to realize. This implies that inaccuracies greater than about 21 percent of the mean ($\pm 3\sigma$ limits) should seldom occur.

Sensitivity Results - The sensitivity analysis technique used in this study is described in Appendix 2. A sample calculation for estimating the variability of S.P. using assumed values of variability for each parameter in the basic equation is given in Appendix 4, page 144, as an example of calculating the mean and variance of nonlinear functions.

Sensitivity analyses performed with the variables modeled as given in Table 1 show that the relative ranking of the most important variables in terms of their influence on data quality varies as a function of the concentration level of S.P. being measured, characteristics of the

Table 2: SYSTEM SIMULATION RESULTS

Computer Run	Input Values			Simulation Results				Remarks
	(S.P.) $\mu g/m^3$	Q_1 m^3/min	T_t min	S.P. $\mu g/m^3$	σ $\mu g/m^3$	Min $\mu g/m^3$	Max $\mu g/m^3$	
1	200	1.7	1440	200	15	170	259	Parameters modeled as shown in Table 1.
2	100	1.7	1440	100	7.2	84	120	
3	200	1.7	1440	200	10	181	241	E(pH), E($\bar{Q}$), and E(Q_r) controlled by using neutral filters and a continuous flow-rate recorder.
4	100	1.7	1440	101	5	89	116	
5	200	1.7	1440	202	8	179	223	Same as run 4 with no delay between collection and analysis.

particulates, and environmental conditions. Figure 3 summarizes the results in terms of the percent change in the average concentration, S.P., when the parameter takes on its -2σ and $+2\sigma$ values. For parameters modeled as uniform distributions, the lower and upper values of their ranges were used as the $\pm 2\sigma$ values.

Figure 3 shows that the effects of four parameters are independent of monitoring conditions as indicated in the box identified as S_1. The importance of errors due to filter surface alkalinity, E(pH), and filter weighing, E(W), depends on the concentration level being measured as shown in boxes S_2, S_3, and S_4; both parameters show an increasing influence on S.P. as the concentration level being measured decreases.

Boxes S_5 and S_6 show that the possible error in calculating the average flow rate, $E(\bar{Q})$, and the possible error due to unequal sampling rates, E[Q(t), S.P.(t)], are greater for particulates that tend to plug the filter resulting in large flow-rate drops over the sampling period.

Assuming that the sampler is calibrated at a temperature, $\bar{T}$, of 21°C and used in the field at different temperatures, boxes S_7, S_8, and S_9 show the error ranges which would result if the average temperature of the sampling site was within the interval of -4°C to 40°C.

Loss of organic matter is a function of time as well as the original organic content of the sample. The error levels given in boxes S_{10} and S_{11} result from the model of E(OM) as given in Table 1 for the delay intervals specified in each box.

Taking a specific example, if data from a particular sampling site show that the average concentration is usually greater than 150 $\mu g/m^3$, the drop in flow rate is usually greater than 20 percent of the initial flow rate, the site temperature is consistently higher than the temperature of calibration, and the delay between collection and analysis is generally greater than 14 days; then, path $S_1 - S_4 - S_6 - S_9 - S_{11}$ is applicable and the possible error sources can be ranked. In this case a listing of the four most important potential error terms in order of decreasing effect would be $E(\bar{Q})$, E[Q(t), S.P.(t)], $E(Q_c)$, and $E(Q_r)$.

It can be seen that for S.P. concentrations less than 75 $\mu g/m^3$ (box S_2) filter surface alkalinity, E(pH), and filter weighing errors, E(W), both become important as the first and third largest sources of possible errors.

See Appendix 2 for directions in evaluating control procedures and/or monitoring strategies using the linear terms of a Taylor's series expansion as an approximate model for the measurement process.

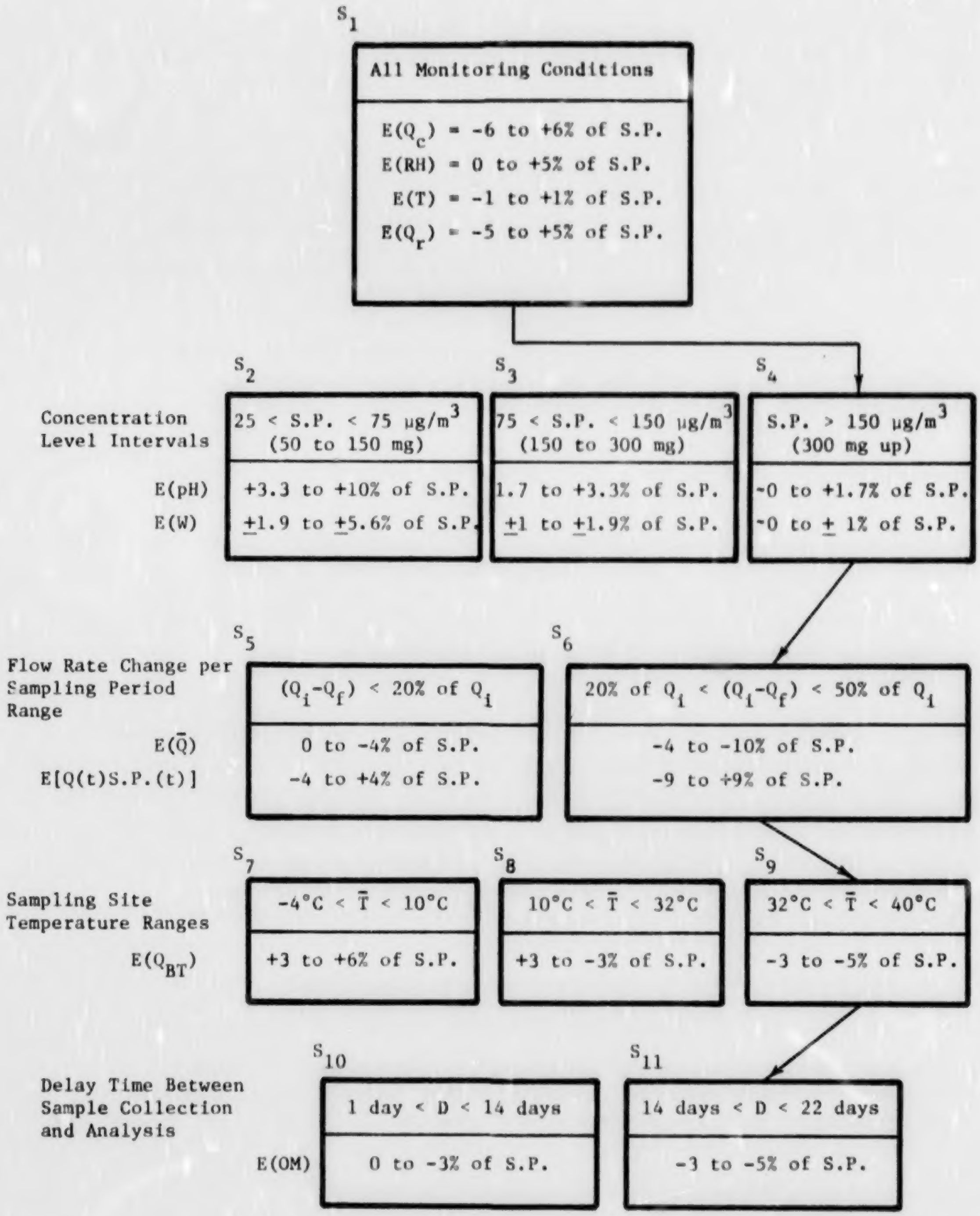


Figure 3: Sensitivity of Measured Concentrations to Various Parameters for Different Sampling Environments

Arriving At Suggested Performance Standards

Use of Collaborative Test Data - The primary documented source of data available for use in providing standards of measurement or baseline data for classifying measurements as either defective or nondefective is the Collaborative study (Ref. 1). A few remarks about these data and the conclusions of the report are given here to aid in relating the use of the data in the corresponding guidelines for development of a quality assurance program.

There was no standardized procedure to guide the laboratories in the use of the High Volume Method, e.g., no standard method for calibration of the High Volume Sampler. The remarks in Ref. 1 concerning the sampling variation relative to location of sources and to ground level indicate the variability of results and hence the difficulty in sampling the "same" atmosphere for checking purposes. This remark is important relative to the use of a mobile reference monitor as discussed in the field document.

The selection of the laboratories and the analysts was on basis of experience with the High Volume Method of sampling suspended particulate matter; thus, the data would be expected to show less variability than if they were based on a random selection of field stations in a sampling network and/or a random selection of the typical operator. Even under these conditions of selection, data from one laboratory were considered too deviant from the others to be included in the final analysis.

The statistical analysis of the data provided estimates of the sources of variation of results within laboratories and among laboratories. A true replicate cannot be accomplished. Data used in the analysis are given in Table 3. The standard deviation of the data for each day varied with the mean, and hence a logarithmic transformation of the data was performed prior to further analysis. This is consistent with the fact that particulate concentrations have been observed to be lognormally distributed. The transformation not only makes the variances of the transformed values homogeneous, but it also yields data which are more nearly normally distributed.

A least squares analysis was performed on the transformed data, resulting in a linear fit to the four data points for each laboratory. As there is no reference value for the concentration on a given day, the mean for all laboratories was used. This value was then used with the observed values for each laboratory in determining the best-fit line to the concentrations for the four days. An analysis of the variation of these data (transformed) indicates that there is no significant variation in the means or slopes of the fitted lines among the 11 remaining laboratories used in the analysis. Thus one infers that the within-day variation among the laboratories is sufficiently large so that the variations in the resulting means and slopes can be explained by chance variation (excluding the outlier laboratory). The fact that there is one laboratory considered an outlier should not be totally ignored in predicting how field monitoring operations will behave. There may be a reasonably small percentage of stations which would behave like the outlier laboratory, that is, until sufficient experience is gained in the use of the techniques.

Table 3. COLLABORATIVE TESTING DATA--HIGH VOLUME METHOD FOR PARTICULATES

Laboratory Number	Particulate Concentration μg/m^3			
	Day 1	Day 2	Day 3	Day 4
222	138	*	87	114
311	125	80	82	113
320	128	72	81	112
341	126	75	83	114
345	127	78	87	124
509	128	74	86	121
572*	128	82	84	112
575*	108	73	72	93
578*	126	77	83	111
600	125	72	80	110
787*	125	76	83	117
799	131	76	86	120
Mean (μg/m^3)	126.3	75.9	82.8	113.4
Standard deviation (μg/m^3)	6.8	3.2	4.1	7.8

*Missing data and recalculated results as described in Ref. 1.

Some specific results of the analysis used in arriving at the suggested performance standards given in the field document are as follows. The relative standard deviation (coefficient of variation) of measurements within a laboratory was estimated as 3 percent; this is referred to as a measure of repeatability. The reproducibility (measure of variation among laboratories) was 3.7 percent. The standard deviation of weight measurements was about 0.8 mg (0.7 to 0.9 mg) for unexposed filters, 1.7 mg for exposed filters. These are considered to be sufficiently different to be explained by other than chance occurrence. The weight losses over 4 days were about 10 mg or 5 percent, assumed to be due to loss of volatile material on the filters. Flow-rate calibration errors were estimated to be 2.1 percent. Intuitively it is felt that these specific values may tend to be on the small side relative to that expected in the field data collection of the monitoring sites.

It is important to note that in comparing two laboratories or two measurements within a laboratory, the variance of the difference is the sum of the variances. In comparing a measurement to a standard or known value, the variance of the difference is the variance of the single measurement (see Appendix 4, page 144 for a more detailed discussion).

Use of Modeling Data - Data from the collaborative test (Ref. 1) are a good indication of the precision obtainable with the High Volume Method. At this time no such collaborative test has been conducted to obtain accuracy data. The development of a performance model and subsequent simulation and sensitivity analyses as discussed previously was in one respect an attempt to generate accuracy data.

The performance model contains several bias terms such as filter surface alkalinity, loss of organic material, ambient temperature, and ambient pressure which would not influence variability in the collaborative test data but are important functions of accuracy.

Simulation and sensitivity data generated from the performance model were used in evaluating alternative monitoring strategies as given in the field document, and from these results performance standards were suggested.

The computer simulation results show the standard deviation varying from approximately 7.5 to 4 percent of the mean value. This is about the same range as that of the mean square error (MSE = $\sqrt{\tau^2 + \sigma^2}$) associated with the 7 alternative strategies given in Table 14 of the field document (Ref. 10). Since it is desirable to encourage high quality data, the lower value of 4 percent of the mean was used in arriving at standards for defining defects. A defect can result from any one of three audit checks. They are 1) flow-rate check, 2) calibration check, and 3) a check of the possible loss of volatile matter between collection and analysis. The suggested performance standards are estimates of what can be achieved under field conditions and yet are restrictive enough to ensure the collection of good quality data. They were not derived from actual data. The suggested performance standards as given in the field document are summarized in Table 4.

All suggested performance standards, control limits, etc., used in the field document are estimated at the 3σ level. If the measurement process is operating at a level resulting in a standard deviation of 4 percent of the actual value, then it would be desirable to define a defect as any measured S.P. concentration deviating more than ± 12 percent from the actual value. A defect is now declared any time one audit check indicates a possible error greater than 9 percent of the actual value (standard deviations for the three checks were all estimated at 3 percent of the mean and were not derived from actual data). If all three checks are in control (i.e., each has a relative standard deviation of 3 percent or less), then the overall standard deviation would be

$$\sigma_T = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \sqrt{27} \approx 5.2.$$

The $3\sigma_T$ value would be 15.6 percent of the mean which is larger but still comparable in magnitude to the desired ± 12 percent given above.

Table 4: **Suggested Performance Standards**

Parameter	Definition for Defining Defects	Suggested Minimum Standards for Audit Rates
1. Flow-Rate Check	$\lvert d_{1j} \rvert > 9$	n=7, N=100; or (n=3, N=15)
2. Calibration Check	$\lvert d_{2j} \rvert > 9$	n=7, N=100; or (n=3, N=15)
3. Elapsed Time Between Collection and Analysis	$\lvert d_{3j} \rvert > 9$	n=7, N=100; or (n=3, N=15)

Parameter	Standards for Corrigible Errors	Suggested Minimum Standards for Audit Rates
4. Filter Surface Alkalinity	$6.5 \leq pH \leq 7.5$	n=7, N=100
5. Weighing of Clean Filters	± 1 mg	n=7, N=100
6. Weighing of Exposed Filters	± 2.7 mg	n=7, N $\geq$ 50; or (n=4, N<50)
7. Data Processing Check	± 3% of S.P.	n=7, N $\geq$ 50; or (n=4, N<50)

Standards for Operation

8. If at any time d[*]=1 is observed (i.e., a defect is observed for either d_{1j}, d_{2j}, or d_{3j}), increase the audit rate to n=20, N=100 or n=6, N=15 until the cause has been determined and corrected.

9. If at any time d=2 is observed (i.e., two defects are observed in the same auditing period), cease operation until the cause has been determined and corrected. When data collection resumes, use an auditing level of n=20, N=100 (or n=6, N=15) until no values greater than ± 6 are observed in three successive audits.

10. If at any time two (2) values of d_{1j}, d_{2j}, or d_{3j} exceeding ± 6 or four[†] values exceeding ± 3 are observed, 1) increase the audit rate to n=20, N=100 or n=6, N=15 for the remainder of the auditing period, 2) perform special checks to identify the trouble area, and 3) take necessary corrective action to reduce error levels.

---

[*] d without a subscript as used here represents the number of defects observed in a sample of size n.

[†] The number three was used in the field document; it should be changed to four.

The extreme possibilities with this method are:

(1) all three audits have values just equal to 9 and of the same sign, and

(2) a defect is declared because of one audit check exceeding 9 when in fact one or both of the other audits are of opposite sign and magnitude resulting in a very small overall error in the measurement.

In the first instance a total error of 27 percent would result and would be indicative of the process being out of control when compared to the desired ±12 percent. The probability of this happening should be very small. The second case shows that good data could be invalidated as the result of the method used to define defects. The probability is again small, but certainly larger than the first probability.

Use of a Mobile High Volume Reference Sampler - An alternative method for checking the precision and accuracy of a field sampler presented for consideration in the field document was by use of a mobile reference sampler. Under the reasonable assumption that two adjacent samplers are sampling approximately the same atmosphere and that the differences in the measurement should result only from instrument and operator errors, the mean and standard deviation of the difference

$$\Delta S.P. = S.P._F - S.P._R$$

would be

$$\text{Mean } (\overline{\Delta S.P.}) = 0 \text{(i.e., both samplers are measuring the same value of C by assumption)}$$

$$\text{Standard deviation } [\sigma(\Delta S.P.)] = \sqrt{\sigma_F^2 + \sigma_R^2} \text{ ; where } \sigma_F \text{ and } \sigma_R$$

are the standard deviations of the measured values for field sampler and reference sampler respectively.

Based on the results of a collaborative test (Ref. 1) showing a relative standard deviation for repeatability of 3.0 percent of the mean value, and assuming that the variability in the data from both samplers is about equal, then

$$\sigma(\Delta S.P.) = \sqrt{(3.0)^2 + (3.0)^2} = 4.2 \text{ percent of the mean value.}$$

A defect, defined as any check in which the difference is equal to or greater than the 3σ value, then would be declared anytime the percent difference, d, is equal to or greater than 12.6 (~13) as computed by

$$d = \frac{\Delta C}{0.5(C_F + C_R)} \times 100 \geq 13.$$

Recommendations

Collaborative Test to Obtain Accuracy Data - In implementing a quality assurance program, it is recommended that a collaborative test be designed and conducted to determine more precisely the accuracy of the High Volume Method. Such a test design is discussed in general terms.

Design of factorial experiment - One means of estimating the effects of various alternatives or strategies on the data quality is to plan an experimental program to assess these effects similar to the manner in which one assesses the measurement errors/variations in a collaborative test. Some key factors in the data quality for suspended particulate matter are the surface pH of the filter, the environment, type of flowmeter (rotameter or continuous recorder), and calibration technique. Other factors could become important for a particular agency (region or site) due to the peculiar environmental effects and their interaction with the operator and/or equipment. The selected factors are used to illustrate a means of studying a set of typical factors.

In order to efficiently study several factors as those above, it is desirable to design an experiment to consider all of the factors simultaneously. Such an experiment is referred to in the literature as a factorial experiment. Figure 4 below indicates how the experimental plan can be approached for the selected factors.

This experiment would require 3×2×2×2 = 24 tests to be made, (i.e., samples to be collected and analyses of the concentration in $\mu g/m^3$ to be determined). This number of tests would seem to be practical; however, it is possible to reduce this number through the judicious selection of the combinations of levels of the several factors. Such an experiment is called a fractional factorial experiment. This fractional approach certainly could be advisable when considering a greater number of factors and/or levels of the factors, resulting in a larger number of tests.

The analysis of the concentration data for the above experiment follows a reasonable standard statistical method known as analysis of variance, in which the total variation in the data is subdivided into meaningful components due to the various factors of interest: environment, flowmeter, pH level, sampler (or calibration error), and the appropriate interactions of these factors. The effect of the sampler will really include all of

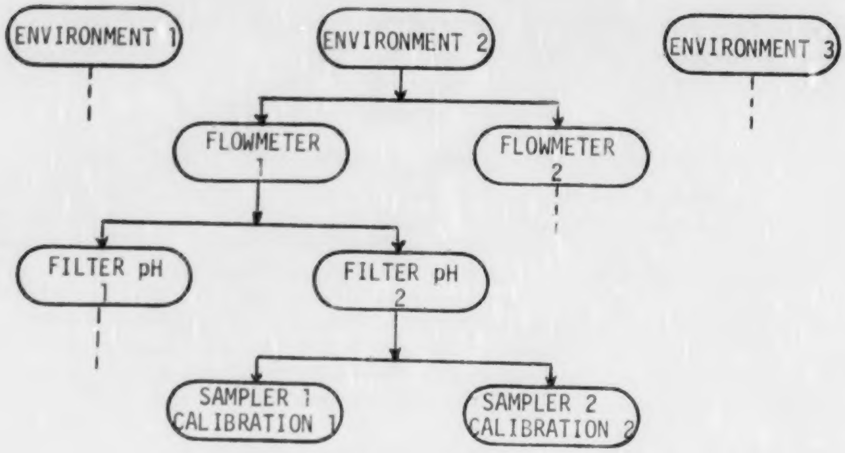


Figure 4: Experimental Plan for Selected Factors

those errors associated with two independently operated and calibrated samplers at the same site. Even if all of the effects could be absolutely controlled, there would be some expected variation in the results from one sampler to another. Although the environments cannot be controlled, it is desirable to select them for varying conditions, e.g., degree of stickiness of the particulates or the percent of volatile matter. This planned environmental effect would then be studied as to how it interacts with or affects the other factors to be studied.

Statistical Analysis of Field Data - Alternative to a collaborative test, audit, data, and other operational data should be evaluated and analyzed in a statistical manner as soon as possible in order to adjust, eliminate, or add checks and standards as given in the field document.

Further Use of the Performance Model - Also, once data from the factorial experiment or a broad spectrum of field data from quality assurance programs are available, they should be used to model the variables in the performance model which in turn could be used to evaluate known and hypothesized alternative monitoring strategies.

REFERENCE METHOD FOR DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

General

The manual (non-continuous) measurement of atmospheric SO_2 by the Pararosaniline Method is subject to the same type errors as the High Volume Method in the previous subsection. Unlike the High Volume Method, however, the Pararosaniline Method from sample collection through sample analysis can be checked by sampling from a test atmosphere containing a known concentration of SO_2. These test atmospheres can be generated with calibrated SO_2 permeation tubes. But permeation tube calibration setups are expensive and sensitive to many parameters which may prove difficult to control in other than well-equipped laboratories. Permeation tubes are not widely used at this time for calibrating the Pararosaniline Method.

Without a permeation tube to generate test atmospheres, a calibration curve is developed by measuring control samples containing a known quantity of sulfite solution. This method only validates the analysis portion of the measurement. It will not in any way serve to detect or indicate errors made in sample collection, sample handling, or performing calculations.

A collaborative test of the method was conducted (Ref. 11) using permeation tubes to generate the test atmospheres. Each participant prepared and measured the test atmosphere in his own laboratory with his own equipment. All sampling periods were 30 minutes.

Results from the collaborative test probably do not reflect the variability characteristic of field conditions and particularly 24-hour sampling periods. Major components of variability which are not accounted for in the collaborative test results include:

(1) Deterioration of SO_2 in the collected sample which is a function of time and temperature.

(2) Sampling system problems such as temporary plugging of the sample line, loss of critical flow, system leaks, or temporary loss of electrical power.

(3) Error in determining and correcting for temperature.

(4) Possible deterioration of SO_2 due to exposure of the sample to direct sunlight during collection or handling.

All the above terms except (3) would act as negative biases in the measured results. The collaborative test results show no bias.

For much the same reasons then, the same approach was used in writing the guidelines for developing a quality assurance program for this method as was used for the High Volume Method. This involved performing a functional analysis of the measurement process which consisted of identification and modeling of important variables and parameters, development of a performance model of the measurement process, and simulation and sensitivity analyses using the performance model was performed as a means of evaluating the overall measurement process. Each phase of the analysis is discussed below.

Identification and Modeling of Important Parameters

Measurement of sulfur dioxide in the ambient atmosphere by the manual pararosaniline method requires a sequence of operations and events that yields as an end result a number that serves to represent the average mass of sulfur dioxide per unit volume of air over the sampling period.

The measurement process can be roughly divided into three phases. They are: 1) sample collection, 2) sample analysis, and 3) data processing.

Sample Collection - The sample collection phase of the measurement method contains numerous sources of error. Some errors can be eliminated by a conscientious operator, while others are inherent and can only be controlled. The sources of error are:

(1) flow-rate calibration,

(2) determination of the volume of air sampled,

(3) elapsed time between sample collection and analysis,

(4) exposure of sample to direct sunlight, and

(5) entrainment or loss of sample other than by evaporation.

Flow-rate calibration - Flow rates for critical orifices calibrated against a wet test meter at different times, by different individuals, but with the same equipment showed a standard deviation of less than 2 percent of the mean (Ref. 12). When a large population of laboratories is considered, the variability would undoubtedly increase significantly. Rotameter calibrations would not be expected to be any more precise than critical orifice calibrations. Large flow-rate calibration errors will be detected by the volume/flow-rate check as part of the auditing procedure. However, to maintain the error at or near a minimum, the wet test meter or calibrated rotameter used for calibration should be checked against a standard such as a good quality soap-bubble meter at least once a quarter.

The flow-rate calibration error was modeled as a normal distribution with a zero mean and a standard deviation of 2 percent of the mean, as

$$E(Q_c) = N(0, 0.02)$$

where $E(Q_c)$ = calibration error expressed as a decimal fraction of the mean value, and

$N(0,0.02)$ = a normal distribution (see Appendix 3 for a discussion of the normal distribution) with a zero mean and a standard deviation of 0.02.

Determination of the volume of air sampled - The volume of air sampled is estimated from the calibration of the critical orifice before and after sampling (flow-rate readings before and after sampling if a rotameter is used) and the sampling period time. Such an estimate assumes that any change in flow rate during the sampling period is linear with time. Nonlinear changes due to such things as temporary plugging of the line or critical orifice from condensed moisture are not detected by this method. Also, a system leak between the absorber and the critical orifice would introduce an error in the calculated volume unless detected and corrected by the operator prior to sample collection.

Errors in the calculated volume are checked as part of the recommended auditing process by using a wet test meter on-site to measure the integrated volume. A discrepancy in the integrated volume measured by the wet test meter and the volume calculated in the usual manner implies that there could be:

(1) system leaks.

(2) nonlinear changes in flow rate,

(3) an error in the sampling period timer, or

(4) flow-rate calibration error.

Each item should be checked and verified.

For the 24-hour sampling period where a calibrated rotameter is used to read the flow rate before, during, and after the sampling period, as part of the auditing process, only items (1) and (4) above can be detected. There is only a small possibility of detecting a temporary change in flow rate (item 2). The timer should be checked independently against an elapsed time indicator at least every six months. Flow-rate calibration error was discussed in the previous subsection.

An additional source of error in estimating the sample volume is due to the inability to determine an average temperature or pressure for the sampling period. This is more important for 24-hour samples than it is for the shorter periods. The recommended method for estimating the average temperature is to average the minimum and maximum temperatures as read from a minimum-maximum thermometer stored in the sampling train box. The average pressure can be taken as the average pressure for that location.

For modeling purposes characteristic distributions and variances were assumed for each of the error sources. The means and standard deviations of the parameters are given in percent. Errors in the sampled volume due to the various parameters are modeled as follows:

(1) System leaks as a uniform distribution (see Appendix 3 for a discussion of the uniform distribution) -0.04 to 0, symbolized by

$$U(-0.04 \text{ to } 0).$$

(2) Intermittent plugging of lines, loss of critical flow, or loss of power were combined and treated as a uniform distribution symbolized by

$$U(-0.04 \text{ to } 0).$$

(3) Inaccuracies in the sampling period timer or elapsed time indicator are not critical for 24-hour samples. For example, a 14-minute error in the sampling period time would represent a 1 percent error in measured concentration. In the computer analyses the error in the sampling period time was simulated with a normal distribution with a zero mean and a standard deviation of 7 minutes. It is symbolized by

$$E(T) = N(0, 7)$$

where $E(T)$ = error in sampling period time in minutes, and

$N(0,7)$ = normal distribution (see Appendix 3) with a zero mean and a standard deviation of 7.

(4) Error in determining the actual average temperature of the sampling period was modeled as a normal distribution with a zero mean and a standard deviation of approximately 1 percent. It is symbolized by

$$N(0, 0.0112).$$

(5) Error in determining the actual average pressure of the sampling period was modeled as a normal distribution with a zero mean and a standard deviation of approximately 0.6 percent. It is symbolized by

$$N(0, 0.0057).$$

Elapsed time between sample collection and analysis - It has been shown that exposure of the sample to temperatures above about 5°C results in SO_2 losses. A loss of about 1.8 percent per day occurred at 25°C and no losses were observed at 5°C (Ref. 13).

It is obvious that long delays between sample collection and analysis at unrefrigerated or uncontrolled conditions will result in sizeable errors.

Every means should be employed to minimize the time that a sample is exposed to temperatures above 5°C.

The difference in the weight of SO_2 collected and the weight of SO_2 in the sample at analysis was modeled as a function of the elapsed time in days and the average temperature of exposure in centigrade. The mathematical model for the loss of SO_2, E(L), is

$$E(L) = \begin{cases} 0; \text{ for } t \le 5°C \\ 0.00059\ (t-5)\ (d); \text{ for } 5°C \le t \le 22°C \\ 0.025\ (t-18)\ (d); \text{ for } 22°C < t < 40°C \end{cases}$$

where E(L) = loss of SO_2 expressed as a decimal fraction of the initial SO_2 content in the collected sample,

t = average exposure temperature in °C and modeled as a normal distribution with a mean of 15°C and a standard deviation of 5°C, written as N(15,5), and

d = the elapsed time in days and is modeled as a uniform distribution with a range of 0 to 10, written as U(0 to 10).

This is essentially a piecewise linear approximation of an exponential function. When more data are available, a simple exponential model probably should be used.

Exposure of sample to direct sunlight - Exposure of the sample to direct sunlight during or after collection can result in deterioration of the sample. Losses from 4 to 5 percent were experienced by samples exposed to bright sunlight for 30 minutes (Ref. 13). Therefore, it is recommended that the bubbler be wrapped in tin foil anytime it is exposed to the direct sunlight for more than the 1 or 2 minutes that might occur in the normal transfer of 24-hour samples from the sampling train box to the shipping block.

This error (see X(15) in Table 5, page 40) was treated as a percent weight loss in the collected SO_2. It was modeled as a uniform distribution ranging from -0.05 to 0, U(-0.05 to 0), which is equivalent to a 0 to 30 minute exposure according to the above referenced data.

Entrainment or loss of sample - Any loss of sample by means other than evaporation during sample collection should result in the operator's invalidating the sample. Large errors can result when:

(1) the operator fails to distinguish between sample loss and evaporation and brings the absorber to the mark with fresh TCM,

(2) the operator fails to bring the absorber to the mark and the laboratory assumes that he did and that part of the sample was lost during shipment, or

(3) the operator is careless in bringing the exposed absorber to the mark (e.g., ± 1 mℓ represents a 2 percent error in a 24-hour sample).

This error (see X(12) in Table 5, page 40, was also treated as a percent weight loss and modeled by a uniform distribution (see Appendix 3) ranging from -0.05 to 0 symbolized by

U(-0.05 to 0).

Sample Analysis - To realize a high level of precision and accuracy from the analysis phase of the method, the analyses must be done carefully, with close attention to temperature, pH of final solution, purity of the chemicals and water, and standardization of the sulfite solution (Ref. 13).

Important parameters in the analysis phase include:

(1) purity of the chemicals and water,

(2) spectrophotometer calibration,

(3) pH of the solution being analyzed,

(4) temperature at analysis compared to temperature at calibration, and

(5) volumetric measurement of exposed sample solution and aliquot.

Purity of chemicals and water - Purity of the chemicals and water is extremely important in obtaining reproducible results because of the high sensitivity of the method. As recommended in the Operations Manual of the field document (Ref. 14), the purity of the pararosaniline dye is checked and, if necessary, purified and assayed before use. The water used must be free of oxidants and should be double-distilled when preparing and protected from the atmosphere when stored and transferred from container to container. All other chemicals should be ACS grade and special care exercised not to contaminate the chemicals while in storage or when removing portions for reagent preparation. Errors due to these sources would appear as excessive variability in the calibration curve, measurement of reagent blanks, and measurement of control and/or reference samples.

Spectrophotometer calibration - The spectrophotometer should be adjusted for 100 percent transmittance when measuring a sample cell of pure water prior to each set of determinations. Also, the calibration of the wavelength scale and the transmittance scale should be checked periodically or any time reference samples cannot be measured within prescribed limits after checks of the reagents and water have failed to identify the trouble.

The calibration of the wavelength scale can be checked by plotting the absorption spectrum (in the visible range) of a didymium glass which has been calibrated by the National Bureau of Standards.

The transmittance scale can be checked using a set of filters from the National Bureau of Standards.

Spectrophotometer calibration error would result from deterioration of the spectrophotometer between calibrations and should be detected by an increased variability in measuring reference and/or control samples.

pH of the solution being analyzed - For this method the maximum sensitivity occurs when the pH of the final solution is in the region of 1.6 ± 0.1 (Ref. 13). If care is exercised in reagent preparation and the analysis phase, the pH should always be within the above limits. Phosphoric acid provides the pH control; hence, it should be checked first when the pH is detected outside the above range.

Temperature - Temperature affects the rate of color formation and fading of the final color. Also, the reagent blank has a very high temperature coefficient. Therefore, it is important that the temperature at analysis be within ± 2°C of the temperature at which the calibration curve was developed. If the normal room temperature varies more than ± 2°C from a set value, it is recommended that a constant-temperature bath be used if a high degree of accuracy is desired.

Large variations in temperature would be detected as errors in measuring control samples and/or reference samples.

All of the above topics are part of the analysis phase of the measurement. Therefore, it appeared feasible to incorporate these error terms into the measured absorbance of a sample through the variability in the slope and intercept of the calibration curve, the error in measuring control samples, and the variability in reagent blank measurements. All values used in the models were taken from collaborative test results (Ref. 11) as explained in the subsection Use of Collaborative Test Data, page 44. The individual models are:

(1) The variability in the slope of the calibration curve was treated as a normal distribution with a mean of 0.03 absorbance unit and a standard deviation of 0.00082 absorbance unit. It is symbolized by

$$N(0.03, 0.00082).$$

(2) The calibration curve intercept on the absorbance axis was modeled as a normal distribution (see Appendix 3 for a discussion of the normal distribution) with a mean of 0.163 absorbance unit and a standard deviation of 0.012 absorbance unit. It is symbolized by

$$N(0.163, 0.012).$$

(3) A standard deviation of 0.4 μg SO_2 for the measurement of control samples multiplied by the average slope (0.03 absorbance unit/μg SO_2) yields a value of 0.012 absorbance unit as the standard deviation for the variability in the measured absorbance of control samples. This was modeled as

$$N(0, 0.012).$$

Measured absorbance, A, of field samples was modeled as

$$A = N(0.03, 0.00082) \times (\mu g\ SO_2) + N(0.163, 0.012) + N(0, 0.012)$$

where μg SO_2 is the actual quantity of SO_2 in the aliquot of the sample.

Volumetric measurements - In the analysis of a 24-hour sample, a 5 mℓ aliquot is taken from the sample for analysis. The results are then multiplied by a dilution factor of 10 to give the SO_2 concentration of the field sample. Treating the dilution factor as

$$10 \pm \sigma = \frac{50 \pm \sigma_1}{5 \pm \sigma_2}$$

means that $\sigma = \sigma_1^2 + \sigma_2^2$. Assuming that $\sigma_1 = \sigma_2 = 2$ percent of the average value, i.e., 5 mℓ ± .1 mℓ and 50 mℓ ± 1 mℓ, then the standard deviation of the dilution factor is 2.8 percent of 10 or $\sigma = 0.28$. The dilution factor was modeled as a normal distribution with a mean of 10 and a standard deviation of 0.28, written as

$$D = N(10, 0.28)$$

where D = a dimensionless dilution factor (normally 10 for a 24-hour sample) including error, and

N(10,0.28) = a normal distribution (see Appendix 3) with a mean of 10 and a standard deviation of 0.28.

Table 5: PARAMETER EFFECTS MODELS AND VARIABLE MODELS AS USED IN SIMULATION AND SENSITIVITY ANALYSES (SO_2)

Parameter Effects Models	Variable Models
$A' = X(1) \times W_a + X(2) + X(3)$	$X(1)$ = slope of calibration curve = $N^*(0.03, 0.00082)$
$A'_o = X(2) + X(4)$	$X(2)$ = Y intercept of calibration curve = $N(0.163, 0.012)$
	$X(3)$ = variability in measuring control samples = $N(0,0.012)$
	$X(4)$ = variability in measuring reagent blanks = $N(0,0.015)$
$D' = N(10, 0.28)$	$X(5)$ = dilution factor = D; = $N(10, 0.28)$
$\bar{Q}_a = \bar{Q}_R\ [1 + X(6) + X(7) + X(8) + X(9) + X(10)]$	$X(6)$ = calibration error = $N(0, 0.02)$
where $\bar{Q}_a$ = actual average flow rate through absorber,	$X(7)$ = air leaks in system = $U^{**}(-0.04 \text{ to } 0)$
	$X(8)$ = line plugging, loss of power, etc. = $U(-0.04 \text{ to } 0)$
and $\bar{Q}_R$ = measured (calculated) flow rate	$X(9)$ = error due to average temperature estimate = $N(0,0.0112)$
	$X(10)$ = error due to average pressure estimate = $N(0, 0.0057)$
$E(T) = N(0, 7)$	$X(11)$ = error in sampling period time = $N(0, 7)$
$W_a = \dfrac{(\mu g\ SO_2/m^3)_T}{V_a}\ [1 + X(12) + E(L) + X(15)]$	$X(12)$ = loss of sample = $U(-0.05 \text{ to } 0)$
	$X(13)$ = average temperature sample is exposed to = $N(15, 5)$
where W_a = actual weight of SO_2 collected,	$X(14)$ = elapsed time between collection and analysis = $U(0 \text{ to } 10)$
$V_a = \bar{Q}_a \times T_a = \bar{Q}_a \times [1440 + X(11)]$,	$X(15)$ = error due to loss of SO_2 from exposure of sample to sunlight = $U(-0.05 \text{ to } 0)$
$(\mu g\ SO_2/m^3)_T$ = true SO_2 concentration, and	
$E(L) = \begin{cases} 0; \text{ for } X(13) \le 5°C, \\ 0.0059(X(13)-5)(d) \text{ for } 5°C < t < 22°C \\ 0.025\ (X(13)-13)(d) \text{ for } 22°C < t < 40°C \end{cases}$	

---

[*] $N(\mu,\sigma)$=normal distribution, see Appendix 3 for a discussion of the normal distribution.

[**] U(a to b)=uniform distribution over the interval ab, see Appendix 3 for a discussion of the uniform distribution.

Data Processing - Data processing, starting with the raw data through the act of recording the measured concentration on the SAROAD form, is subject to many types of errors. The approach recommended in the Operations Manual of the field document means that one can be about 55 percent confident that no more than 10 percent of the reported concentrations are in error by more than ± 3 percent due to data processing error.

The magnitude of data processing errors can be estimated from, and controlled by, the auditing program through performance of periodic checks and making corrections when large errors are detected. A procedure for estimating the bias and standard deviation of data processing errors is given in the Management Manual of the field document. It was not included in the performance model for this analysis, but could be included in the overall assessment of data quality as explained in the field document.

All parameter and variable models are summarized in Table 5.

Performance Model

A mathematical performance model of the measurement process incorporating the error terms discussed in the previous subsection and summarized in Table 5 was derived from the basic equation

$$(\mu g\ SO_2/m^3)_m = \frac{(A - A_o)\ (10^3)\ (B_s)}{V} \times D$$

where $(\mu g\ SO_2/m^3)_m$ = measured SO_2 concentration,

A = absorbance of sample,

A_o = absorbance of reagent blank,

B_s = reciprocal of the calibration curve slope,

D = dilution factor, and

V = volume of air sampled.

Introducing the error terms gives

$$(\mu g\ SO_2/m^3)_m = \frac{(A' - A_o')\ (B_s)\ (D')}{\bar{Q}_R \times T}$$

where A', A_o', and D' are defined in Table 4,

B_s = reciprocal of the calibration curve slope,

$\bar{Q}_R$ = correct average flow rate, and

T = correct sampling period time.

Simulation Results - Simulation and sensitivity data were obtained for 3 levels of SO_2 concentration; namely, 800, 400, and 100 μg SO_2/m^3. These data are believed to be representative of the results obtainable from a large population of sampling sites. The parameter and variable models are as given in Table 5. Simulation results are summarized in Table 6. The results show a negative bias of about 5 percent of the true value, i.e., the average measured SO_2 concentration. Standard deviations of 8.9, 9.5, and 24 percent of the true SO_2 concentrations were obtained for concentration levels of 800, 400, and 100 μg/m^3, respectively.

Table 6: SYSTEM SIMULATION RESULTS (SO_2)

Computer Run	Input Values			Simulation Results			
	$\bar{Y}$ (true) μg SO_2/m^3	$\bar{Q}$ cm^3/min	T_t Min	$\bar{Y}$ (measured) μg SO_2/m^3	σ μg SO_2/m^3	Min μg/m^3	Max μg/m^3
1	800	200	1440	757	71	589	954
2	400	200	1440	376	38	290	491
3	100	200	1440	92	24	30	148

Sensitivity Results - The sensitivity analysis technique used in this study is described in Appendix 2. A sample calculation for estimating the mean and variance of nonlinear functions is presented in Appendix 4, page 154, using suspended particulates as an example.

Table 7 summarizes the sensitivity results by listing the 5 variables effecting the largest variability in the measured data for each of the three levels of SO_2 concentration.

In all three runs the 5 most significant variables accounted for the major portion of data variability (see Appendix 2). The order of ranking changed significantly from one concentration level to another for the 5 most important variables while the order of the 8 variables having the least effect did not change.

See Appendix 2 for directions in evaluating control procedures and/or monitoring strategies using the linear terms of a Taylor's series expansion as an approximate model for the measurement process.

Table 7: Sensitivity Analysis Results (SO_2)

Computer Run	Input Values			Variables in Decending Order of Importance														
	Y (true)	$\bar{Q}$ cm^3/min	T_t min	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	800	200	1440	X13	X5	X1	X4	X6	X14	X3	X15	X12	X9	X7	X8	X10	X11	X2
2	400	200	1440	X13	X4	X3	X5	X1	X6	X14	X15	X12	X9	X7	X8	X10	X11	X2
3	100	200	1440	X4	X3	X13	X5	X1	X6	X14	X15	X12	X9	X7	X8	X10	X11	X2

Variables appearing in the top 5 for each run are:

X13 = Average temperature that the sample is exposed to between collection and analysis.
X4 = Variability in measuring reagent blanks.
X5 = Dilution factor.
X3 = Variability in measuring control samples.
X1 = Slope of calibration curve.
X6 = Calibration error.

Arriving at Suggested Performance Standards

The suggested performance standards as given in the field document are reproduced in Table 8.

The standard proposed for the volume/flow-rate check was not based on actual data. The equipment error (i.e., wet test meter, rotameter or critical orifice) should be small in comparison to system errors such as leaks, lines plugging, etc. It was felt that the difference in the volume as measured by the audit and that calculated in the normal manner could not be lower than 3 percent of the audited value on the average. A relative standard deviation of 3 percent was assumed, making the value for defining a defect 9(3σ).

Also, the suggested standard for correcting data processing errors was an estimate. It was desired to use a standard large enough to accept round-off errors yet small enough to prevent data processing errors from becoming a major component of overall system error.

Use of Collaborative Test Data - The results of a collaborative test of the manual pararosaniline method of analyzing SO_2 are given in Reference 11. Some specific remarks concerning this study and its use for the field document on SO_2 are contained herein. Fourteen laboratories provided the results, based on analysis of pure synthetic atmospheres using the 30-minute sampling procedure and the sulfite calibration method. The permeation tubes provided the method of generating test atmospheres. It should be remarked that this procedure was not recommended in the guideline document because of its high cost. Furthermore, it appears that there may have been considerable difficulty in the generation/measurement of the test atmospheres resulting in a large variability between labs (or large reproducibility measures). Seventeen (17) labs participated in the first test. Based on the results of this test and the lack of uniformity in test procedure, this test was voided and a second test program with 14 participants was run. A familiarization session of three days in length was conducted between the first and second tests. This experience in running the test twice indicates that until the field laboratories gain the needed experience, their results can be subject to large variability and inaccuracies.

The procedure required that a separate calibration curve be prepared for each day. Also certain reagents were to be made fresh each day. The concentrations were nominally 150, 275, and 820 $\mu g\ SO_2/m^3$. A total of 378 determinations were made (14 labs × 3 concentrations × 3 days × 3 replications = 378). Two cases of atypical data were noted and corrected by an appropriate statistical procedure. Four instances of inadvertent errors in arithmetic operations were noted and corrected. (A checking procedure should uncover such errors in routine field operations, e.g., such as the procedure suggested in the field document).

Table 8. SUGGESTED PERFORMANCE STANDARDS

Standards for Defining Defects

1. Volume/Flow-Rate Check; $|d_{1j}| \geq 9$

2. Measurement of Reference Samples; $|d_{2j}| \geq 1.2\ \mu g\ SO_2$[*]

Standard for Correcting Data Processing Errors

3. Data Processing Check; $|d_{3j}| \geq 3$

Standards for Audit Rates

4. Suggested minimum auditing rates for data error; number of audits, n = 7; lot size, N = 100; allowable number of defects per lot, d = 0.

Standards for Operation

5. If at any time d = 1 is observed (i.e., a defect is observed) for either d_{1j} or d_{2j}, increase the audit rate to n = 20, N = 100 until the cause has been determined and corrected.

6. If at any time d = 2 is observed (i.e., two defects are observed in the same auditing period), stop collecting data until the cause has been determined and corrected. When data collection resumes, use an auditing level of n = 20, N = 100 until no defects are observed in three successive audits.

7. If at any time either one of the two conditions listed below is observed, 1) increase the audit rate to n = 20, N = 100 for the remainder of the auditing period, 2) perform special checks to identify the trouble area, and 3) take necessary corrective action to reduce error levels. The two conditions are:

   (a) two (2) d_{1j} values exceeding ± 6, or
   four[†] (4) d_{1j} values exceeding ± 3.

   (b) two (2) d_{2j} values exceeding $\pm 0.8\ \mu g\ SO_2$, or
   four[†] (4) d_{2j} values exceeding $\pm 0.4\ \mu g\ SO_2$.

---

[*] $\mu g\ SO_2$ = total quantity of SO_2 in the reference sample.

[†] A value of three was used in the field document; it should be changed to four.

Sources of error – Forty-two independent calibrations of absorbance vs. concentration were run (14 labs × 3 days). From these calibrations the mean standard error of estimate, mean slope and its standard deviation, mean intercept and its standard deviation were determined.

Mean std. error of estimate = 0.010 absorbance units.

$$\text{Mean slope} \pm \text{std. dev.:}\quad 0.03 \pm \begin{cases} 0.00082 & \text{absorbance units}/\mu\text{g } SO_2 \text{ (within lab)} \\ 0.00195 & \text{absorbance units}/\mu\text{g } SO_2 \text{ (among lab)} \end{cases}$$

$$\text{Mean intercept} \pm \text{std. dev.:}\quad 0.163 \pm \begin{cases} 0.013 & \text{absorbance units}/\mu\text{g } SO_2 \text{ (within lab)} \\ 0.022 & \text{absorbance units}/\mu\text{g } SO_2 \text{ (among lab)} \end{cases}$$

The variation within laboratories was used as a measure of precision in modeling the variation of the measured concentration of SO_2. See the subsection on Identification and Modeling of Important Parameters, page 33. The upper end of the calibration curve contributed most heavily to the standard error of estimate, suggesting the errors of measurement to be dependent on concentration. The variation between laboratories was significant relative to the variation between days, which is expected as the reagent blank is temperature sensitive.

In the field document the acceptable range in the slope and intercept of the calibration curve was given as the mean ± 3σ values stated above.

Thirty-six control samples (12 labs × 3 days) were analyzed for the differences between the amount of SO_2 taken and the amount found, both in micrograms. The standard deviation for within-laboratory variation was 0.40 μg (24 D.F.)* and for between-laboratory variation was 0.55 μg (11 D.F.). Thus the absolute differences should fall below

1.2 μg, 99.7 percent of the time for within a laboratory and

1.65 μg, 99.7 percent of the time for between laboratories.

These limits provide a method of using the results of the control sample. If the deviation exceeds the value given above, appropriate checks should be made.

The standard for measuring reference/control samples when prepared from sulfite solution was suggested as 1.2 μg SO_2 (3σ level).

---

*D.F. = degrees of freedom associated with the estimated standard deviation.

Replicate, repeatability, and reproducibility error – The errors of replication (within day-lab), repeatability (between day-within lab), and reproducibility (between lab) are given below for convenient reference. All errors are assumed to be linear functions of the concentration of SO_2 (denoted by Y) and are expressed in $\mu g\ SO_2/m^3$.

$$\sigma(\text{replication}) = 10(0.7 + 0.001Y)\mu g\ SO_2/m^3$$

$$\sigma(\text{repeatability}) = 21(0.7 + 0.001Y)\mu g\ SO_2/m^3$$

$$\sigma(\text{reproducibility}) = 41(0.7 + 0.001Y)\mu g\ SO_2/m^3$$

In order to compare differences in two measurements corresponding to the above, each of the above must be multiplied by $\sqrt{2}$ (1.96) = 2.77, for including 95 percent of the differences, and $\sqrt{2}$ (3) = 4.23 for including 99.7 percent of the differences.

Statistical analysis – The dependence of the variability of the measurements on the concentration suggest a transformation of all of the data prior to refined analysis. A logarithmic transformation of Y (observed concentration) was performed,

$$Z = 1000 \log_e(7 + 0.01Y) .$$

A linear effect model was employed, i.e., the observed concentration was assumed to be equal to an overall mean + a lab effect + material (concentration) effect + day effect + lab × material interaction effect + day × material interaction effect + replicate effect. All the effects were found to be significant.

In all, three methods of statistical analysis were performed, and the resulting estimates of replication, repeatability, and reproducibility errors were plotted. (See Fig. B-3 in Reference 11.) These results were in good agreement.

No statistically significant systematic error or bias was determined from these data.

One should refer to Reference 11 for the detailed data and results.

Recommendations

The auditing scheme as given in the field document does not effectively check the overall data quality unless the other special checks are made and the suggested control limits are adhered to.

An improvement in the auditing scheme would be realized if equipment were available to measure directly the sampled volume without invalidating the sample and at flow rates as low as 200 cm^3/min as is required for the 24-hour sampling period.

Data derived from the implementation of a quality assurance program should be analyzed in a statistical manner as soon as practical to readjust, eliminate, or add to the special checks and control limits suggested in the field document.

Also, the performance model should be verified using data of known quality from the quality assurance program and used for evaluating known and hypothesized alternative monitoring strategies.

REFERENCE METHOD FOR THE MEASUREMENT OF PHOTOCHEMICAL OXIDANTS CORRECTED FOR INTERFERENCES DUE TO NITROGEN OXIDES AND SULFUR DIOXIDE

General

The reference method for the continuous measurement of ozone utilizes the chemiluminescent reaction between ethylene gas and ozone. Chemiluminescence produced by mixing ethylene gas and the sample air is detected and amplified by a photomultiplier tube. The magnitude of the chemiluminescent signal is proportional to the ozone concentration.

Other pollutants normally found in ambient air do not interfere with this reaction. Therefore, the total measurement process can be audited with two checks. The two checks are 1) measurement of control samples, i.e., measuring samples of known O_3 concentrations, and 2) data processing checks. Results from the two audit checks are used to estimate the quality of the reported data.

Arriving at Suggested Performance Standards

The performance standards suggested in the field document (Ref. 15) are given in Table 9. Each standard is discussed individually.

Standard for Defining Defects - A defect is declared anytime a control sample cannot be measured within $\pm(0.01 + 0.075 \times \text{ppm } O_3)$ of its true value. This standard was derived from a provisional report of a collaborative test of the reference method (Ref. 16) and from the analysis of a set of data provided by EPA. Specifically, the collaborative test results showed a standard deviation of $(0.009 + 0.07 \times \text{ppm } O_3)$ for repeated measurements of the same sample by a given laboratory. The standard for defining a defect is given at the 3σ level. This can be changed as the manager or supervisor wishes. Analysis of the ozone data as given below shows that if the coefficient of variations as computed in Tables 14 and 16 (pages 54 and 55) are combined to arrive at a standard deviation for measuring control samples, agreement with the standard deviation from the collaborative test is very good.

Analysis of Ozone Data - Calibration data for a chemiluminescent ozone monitor were taken on three days--October 31, November 1, and November 15, 1972--for six sleeve setting positions of the ozone generator. Three repeat or replicate measurements were taken each day (with the exception of one day for one sleeve setting) with both the KI Method and the monitor. These data were provided to RTI by EPA for analysis and used in developing standards for the ozone field document. The data are given below, followed by an analysis of the variability of the measurements between and within days (replicates) for each method.

Table 9. SUGGESTED PERFORMANCE STANDARDS

Standards for Defining Defects

1. Measurement of Control Samples; $d_{1j} \geq \pm (0.01 + 0.075 \times \text{ppm}$[*] $O_3)$

Standard for Correcting Data Processing Errors

2. Data Processing Check; $|d_{2j}| \geq 0.01$ ppm

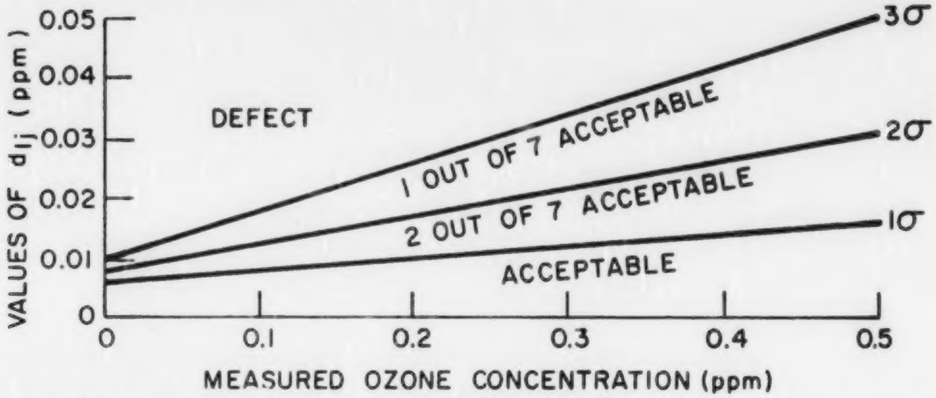
Standards for Audit Rates

3. Suggested minimum auditing rates for data error; number of audits, n=7; lot size, N=100; allowable number of defects per lot, d[+] = 0.

4. Suggested minimum auditing rates for data processing error; number of audits, n=2; lot size, N=24; allowable number of defects (i.e., $|d_{2j}| \geq 0.01$ ppm) per lot, d=0.

Standards for Operation

5. Plot the values of d_{1j} on the graph below.



6. If at any time during the auditing period
   (a) one defect (d=1) is observed (i.e., a plotted value of d_{1j} is in the defect region of the graph),
   (b) two plotted points fall in the region between the 2σ and 3σ lines, or
   (c) four d_{1j} values fall outside the acceptable region,

   increase the audit rate to n=20, N=100 until the cause has been determined and corrected.

7. If at any time two defects (d=2) are observed (i.e., two d_{1j} values plot in the defect region in the same auditing period), stop collecting data until the cause has been determined and corrected.

---

[*] ppm O_3 is the concentration measured by the KI Method or the output of the ozone generator.

[+] An unsubscripted d represents the number of defects observed from n audits of a lot size of N.

Table 10. CALIBRATION DATA FOR CHEMILUMINESCENT OZONE MONITOR

Sleeve Setting	10-30-72		11-1-72		11-15-72	
	ppm O_3 (KI)	Monitor DC μa	ppm O_3 (KI)	Monitor DC μa	ppm O_3 (KI)	Monitor DC μa
0	0.000	0.0	0.000	0.0	.000	0.0
10	.039	7.5	.040	7.8	.043	8.3
10	.033	7.7	.045	8.0	.048	8.4
10	.035	7.7	.045	8.0	.047	8.5
20	.086	17.3	.097	17.7	.096	19.0
20	.088	17.3	.094	17.6	.096	18.9
20	.085	17.7	.094	17.6	.097	18.9
30	.140	27.5	.149	27.8	.148	29.3
30	.139	27.5	.150	27.9	.149	29.3
30	.141	28.0	.155	27.9	.145	29.2
50	.254	47.5	.259	47.0	.253	49.4
50	.255	48.0	.257	47.0	.251	49.6
50	.249	47.8	.258	46.9	.251	49.5
75	.369	71.5	.384	70.2	.386	74.3
75	.360	71.5	.384	70.2	.383	74.3
75			.388	70.3	.379	74.6

An analysis was first performed for all of the data, excluding the zero sleeve setting and the third measurement in each set of three measurements. This last omission was for convenience of having the same number of measurements for each day for each method. It is obvious from examination of the data that the within-day variation is small compared to day-to-day variation. It is recognized that the within- and between-day variations depend on the sleeve setting, i.e., the ozone concentrations; and thus a second set of analyses were performed for each sleeve setting. In this second set of analyses all of the data were used. The results of both sets of analyses are tabulated below along with some summary remarks.

The analysis of variance, Tables 11 and 12, subdivides the total variation of the data into its component parts. There are 30 measurements, and the total variation corrected for the mean has an associated 30-1=29 degrees of freedom. Similarly, there are 3-days, 3-1=2 d.f., 5 levels, 4 d.f. The interaction D × L is a measure of the variation of the measurements from day-to-day for each level. If for example, the day-to-day effects were identical for each level in that the same added effect could be used for each day for all levels to obtain the results on the next day, this interaction effect would be zero or negligible. However, it is expected that the effect will

Table 11. ANALYSIS OF VARIANCE OF KI DATA

Source of Variation (SV)	Sum of Squares (SS)×10^6	Degrees of Freedom (DF)	Mean Square (MS)×10^6
Total (Corrected for Mean)	436,783.5	29	15,061.5
Days (D)	578.4	2	289.2
Levels (L)	435,808.3	4	108,952.1
D × L	296.3	8	37.0
Replicates Within Days	100.5	15	6.7

Table 12. ANALYSIS OF VARIANCE OF CHEMILUMINESCENT MONITOR DATA

SV	SS × 10^2	DF	MS × 10^2
Total (Corrected for Mean)	1,567,899.4	29	54,065.5
Days (D)	2,316.1	2	1,158.1
Levels (L)	1,564,740.9	4	391,185.2
D × L	821.9	8	102.7
Residual	20.5	15	1.35

be different depending upon the magnitude of the measurement and hence the second set of analyses. The replicates within days have 15 d.f., one for each set of duplicates. The relative magnitudes of the mean squares (obtained by dividing the sum of squares by the respective d.f.) are as expected. Obviously the effect of levels was expected and is really not of interest. The effect of day exceeds that of D × L, and the variation between duplicates is very small.

The second set of analyses, Tables 13 and 15, provides estimates of the variation between and within days, and then total for each sleeve setting and for each method. The expected mean squares, E(M.S.), are derived and used to obtain the estimates of the two components of variation. In each case, except for sleeve setting 75, there were 9 measurements, three days and three within each day. Thus, there are 8 d.f. for total corrected for the mean (not shown in table), 2 d.f. for between days, and 6 d.f. for within days. In all cases the between-day variation was considerably larger than the within-day variation. The results are summarized in Tables 14 and 16 giving the estimate of the total variance of measurements made on an arbitrary day, i.e., the between-day effect and the within-day effect are added to obtain an estimate of repeatability. The total variance $\hat{\sigma}^2$ is given in the first row, followed by the standard deviation $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$, the mean level (ppm O_3 for the KI method and μa for the chemiluminescent ozone monitor), and lastly the coefficient of variation, 100 $\hat{\sigma}/x$, in %. The coefficient of variation enables a direct comparison of data in different units.

One can readily observe that for the KI method, $\hat{\sigma}$ is almost constant over the sleeve settings (ranges of ppm O_3), and hence the coefficient of variation decreases from 14.1 to 2.9% (1.5% at 50 sleeve setting). On the other hand, for the ozone analyzer data, $\hat{\sigma}$ increases steadily from 0.40 to 2.12 μa DC, and the coefficient of variation is reasonably constant, ranging from 5.0% at sleeve setting 10 to 3.1% at 75 sleeve setting (2.7 at 50). If these results can be found to be consistent with results at field monitoring stations, it would suggest restricting the use of the KI Method to high concentrations (above about 0.1 ppm O_3) and using the ozone generator calibration curve (i.e., the average calibration curve as discussed in the field document) for low concentrations when calibrating analyzers. Further data would be needed to check out this inference. In any case the coefficients of variation of the two measurements methods are about the same order of magnitude for sleeve settings 20, 30, 50 and 75. These estimates can be used for standards until additional field data are obtained.

Standard for Correcting Data Processing Errors - In the field document it was assumed that the largest source of error in data processing was in reading the hourly average from a strip chart recording. For a dynamic strip chart trace, estimating the hourly average can be highly subjective. The standard, 0.01 ppm, represents 2 percent of full scale for an analyzer range of 0 to 0.5 ppm. This, in turn, represents two divisions on the strip chart trace. This standard is no more than an estimate.

Table 13. ANALYSIS OF VARIANCE OF KI DATA FOR EACH SLEEVE SETTING

SV	$SS \times 10^6$	DF	$MS \times 10^6$	E (MS)	Estimates ($\times 10^6$)
Sleeve Setting 10					
Days	172.7	2	86.3	$\sigma_W^2 + 3\sigma_B^2$	$\hat{\sigma}_B^2 = 26.0$
Within Days	49.3	6	8.2	σ_W^2	$\hat{\sigma}_W^2 = 8.2$
Sleeve Setting 20					
Days	176.9	2	88.4	Same	= 28.8
Within Days	11.2	6	1.9		= 1.9
Sleeve Setting 30					
Days	198.3	2	99.1	Same	= 31.3
Within Days	31.3	6	5.2		= 5.2
Sleeve Setting 50					
Days	69.7	2	34.8	Same	= 10.2
Within Days	25.3	6	4.2		= 4.2
Sleeve Setting 75					
Days	581.2	2	290.6	$\sigma_W^2 + 2.625\, \sigma_B^2$	$\hat{\sigma}_B^2 = 104.9$
Within Days	75.7	5	15.14	σ_W^2	$\hat{\sigma}_W^2 = 15.14$

Table 14. ESTIMATE OF STANDARD DEVIATION FOR REPEATABILITY $(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)^{1/2}$

	Sleeve Setting				
	10	20	30	50	75
$\hat{\sigma}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2 (ppmO_3)^2$	34.2×10^{-6}	30.7×10^{-6}	36.5×10^{-6}	14.4×10^{-6}	120.0×10^{-6}
$\hat{\sigma}$ (ppm O_3) =	5.85×10^{-3}	5.54×10^{-3}	6.04×10^{-3}	3.8×10^{-3}	11.0×10^{-3}
Mean = $\bar{x}$ ($ppmO_3$) =	41.7×10^{-3}	92.6×10^{-3}	146.2×10^{-3}	254.1×10^{-3}	379.1×10^{-3}
100 $\hat{\sigma}/\bar{x}$ (%) =	14.1	6.0	4.1	1.5	2.9

Table 15. ANALYSIS OF VARIANCE OF CHEMILUMINESCENT OZONE MONITOR DATA FOR EACH SLEEVE SETTING

SV	$SS \times 10^2$	DF	$MS \times 10^2$	E (MS)	$\text{Estimate} \times 10^2$
Sleeve Setting 10					
Days	89.6	2	44.8	$\sigma_W^2 + 3\sigma_B^2$	$\hat{\sigma}_B^2 = 14.5$
Within Days	7.3	6	1.2	σ_W^2	$\hat{\sigma}_W^2 = 1.2$
Sleeve Setting 20					
Days	398	2	199	Same	65.7
Within Days	12	6	2	Same	2.0
Sleeve Setting 30					
Days	456	2	228	Same	75
Within Days	18	6	3	Same	3
Sleeve Setting 50					
Days	1006.3	2	503.1	Same	166.85
Within Days	15.3	6	2.55	Same	2.55
Sleeve Setting 75					
Days	2568.8	2	1284.4	$\sigma_W^2 + 2.625\sigma_B^2$	$\hat{\sigma}_B^2 = 489.2$
Within Days	0.7	5	0.14	σ_W^2	$\hat{\sigma}_W^2 = 0.14$

Table 16. ESTIMATE OF STANDARD DEVIATION FOR REPEATABILITY $(\hat{\sigma}_B^2 + \hat{\sigma}_W^2)^{1/2}$

	Sleeve Setting				
	10	20	30	50	75
$\hat{\sigma}^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2 (\mu a)^2$	5.7×10^{-2}	67.7×10^{-2}	78.0×10^{-2}	169.4×10^{-2}	489.4×10^{-2}
$\hat{\sigma}(\mu a) =$	.40	.82	.88	1.30	2.21
$\text{Mean} = \bar{x}\ (\mu a) =$	8.0	18.0	28.3	48.1	72.1
$100\ \hat{\sigma}/\bar{x}\ (\%) =$	5.0	4.6	3.1	2.7	3.1

Recommendations

When the collaborative test results (Ref. 17) are published, they should be studied and adjustments made, if deemed necessary, in the standards and/or limits given in the field document.

Data derived from the implementation of a quality assurance program should be analyzed in a statistical manner as soon as possible so as to readjust, eliminate, or add to the special checks and control limits given in the field document. One area needing further work involves the method of calibrating the ozone generator. A primary calibration using the KI method to monitor the output of an ozone generator is recommended in the reference method. The calibration data provided by EPA and presented in the previous subsection are data taken by experienced personnel in the laboratory under reasonably well controlled conditions. These data yield coefficients of variation of 14 percent and 6 percent at ozone concentration levels of about 0.04 ppm and 0.09 ppm, respectively. Under field conditions and with less experienced personnel the variability would be expected to increase significantly.

An ozone generator calibrated against the primary standard (i.e., the buffered KI method) is being supported as a secondary standard for calibrating ozone monitors (Ref. 17). Analysis of the set of calibration data presented previously indicates that for ozone concentrations less than about 0.1 ppm the coefficient of variation (Table 14) for the KI method is larger than the coefficient of variation (Table 16) of the generator output as monitored by a chemiluminescent analyzer. If these data can be verified on a larger scale, it would appear that developing an average calibration curve for an ozone generator from at least 5 primary calibrations, then using the generator as a secondary standard for calibrating the monitors would increase the precision over that obtained by performing a single primary calibration of the analyzer.

TENTATIVE METHOD FOR THE CONTINUOUS MEASUREMENT OF NITROGEN DIOXIDE (CHEMILUMINESCENT)

General

A nitrogen dioxide analyzer operating on the chemiluminescent principle contains a photomultiplier tube for measuring the light energy resulting from the chemiluminescent reaction of nitric oxide (NO) and ozone (O_3).

The measurement cycle begins with a direct measure of NO in the sample for a short period of time. During the second part of the cycle, total nitrogen oxides (NO_x) is determined by first converting NO_2 in the sample air to NO and measuring total NO. Subtracting the original value for NO from the NO_x value gives an estimate of the NO_2 concentration for that measurement cycle (Ref. 18).

Overall data quality from this measurement method can be evaluated with two special checks. They are: 1) measurement of control samples and 2) data processing checks (Ref. 19).

Arriving at Suggested Performance Standards

Suggested performance standards given in the field document are reproduced in Table 17. This method has not been collaboratively tested. The following RTI in-house data were used to arrive at a suggested performance standard for measuring control samples.

A limited amount of precision data were obtained for two NO_2 analyzers from different manufacturers. They will be referred to as analyzer A and analyzer B in the following discussion. A series of zero and span determinations was obtained. The data are given in Tables 18 and 19. These data were analyzed to determine the variation between days and within days (three replicates or voltage readings were taken at each setting or level of NO_2 concentration). All of the analyses are in volts. The results are converted to ppm in the final stage of the analysis.

Table 20 contains an analysis of analyzer A data for 0.0, 0.10, and 0.40 ppm NO_2. The total variation of the analyzer readings corrected for the mean is subdivided into variation between days and within days. The sum of squares (S.S.), degrees of freedom (D.F.) and mean square (M.S.) are given for each source. From these values of the mean squares, one can estimate the variance of the analyzer readings within and between days. These estimates are summarized in Table 22. Tables 21 and 23 contain similar analyses for analyzer B.

The summary Tables, Nos. 22 and 23, give an estimate of the within-day standard deviations. (That among the three readings at one setting or level of NO_2 concentrations, $\hat{\sigma}_W$, the estimate of the between day standard deviation, $\hat{\sigma}_B$, the total $\hat{\sigma} = \sqrt{\hat{\sigma}_B^2 + \hat{\sigma}_W^2}$, the mean analyzer reading, $\overline{X}$, and the coefficient of variation, C.V. The $\hat{\sigma}$'s can be converted from volts to ppm by dividing each value by 2 (i.e., the relative variation in scales).

One readily observes that for both analyzers $\hat{\sigma}$ increases with the level of NO_2 concentration; this is the primary reason for breaking down the analysis into each level. However, the coefficient of variation (ratio of the estimated standard deviation to the mean expressed in percent) does not appear to be constant. In fact, it appears to decrease. The $\hat{\sigma}$'s are plotted as a function of the true concentration levels as shown in Figure 5.

There is a considerable variation between the precisions of the two analyzers. In order to obtain some tentative standards, the two values (one for each analyzer) were averaged. A line was fitted to the resulting averages, and the following equation relating the precision to concentration was derived:

$$\hat{\sigma} = (0.003 + 0.015C_{NO_2})\ \text{ppm}\ .$$

The coefficients are one-half of those given in Figure 5 in order to relate the variation in ppm to that expressed in volts.

The suggested performance standard for measuring control samples is given in Table 17.

Recommendations

This is a relatively new method for measuring NO_2. It has not been subjected to a collaborative test. This should be the most logical step in evaluating the method.

An integrating flask in the sampling line of sufficient size to integrate NO spikes of duration less than or equal to the cycle period should be made standard equipment.

As an alternative to installing an integrating flask, a study should be made of the best way to treat negative values of NO_2 when computing hourly averages. Presently there is no recommended method. In some instances, all negative values are dropped (i.e., treated as zero). This would generally result in a higher-than-true measured hourly average concentration.

Table 17. SUGGESTED PERFORMANCE STANDARDS

Standards for Defining Defects

1. Measurement of Control Samples; $d_{1j} \geq \pm (0.009 + 0.045 \times (C_{NO_2})^{*}_{oj})$ ppm

Standard for Correcting Data Processing Errors

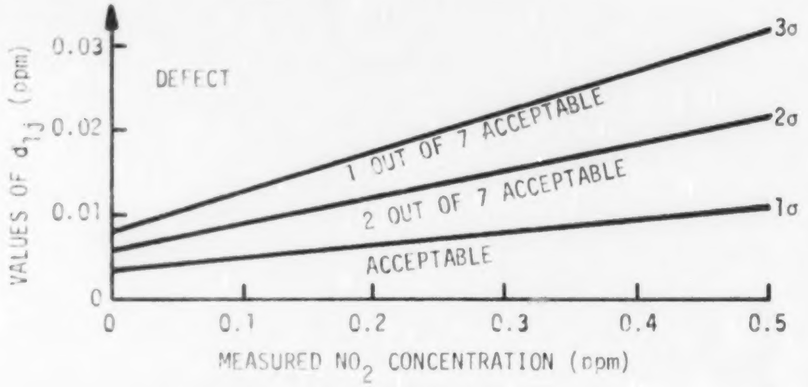
2. Data Processing Check; $|d_{2j}| \geq 0.01$ ppm

Standards for Audit Rates

3. Suggested minimum auditing rates for data error; number of audits, n=7; lot size, N=100; allowable number of defects per lot, $d^{+} = 0$.
4. Suggested minimum auditing rates for data processing error; number of audits, n=2; lot size, N=24; allowable number of defects (i.e., $|d_{2j}| \geq 0.02$ ppm) per lot, d=0.

Standards for Operation

5. Plot the values of d_{1j} on the graph below.



6. If at any time during the auditing period
   (a) one defect (d=1) is observed (i.e., a plotted value of d_{1j} is in the defect region of the graph),
   (b) two plotted points fall in the region between the 2σ and 3σ lines, or
   (c) four d_{1j} values fall outside the acceptable region,

   increase the audit rate to n=20, N=100 until the cause has been determined and corrected.
7. If at any time two defects (d=2) are observed (i.e., two d_{1j} values plot in the defect region in the same auditing period), stop collecting data until the cause has been determined and corrected.

---

*$(C_{NO_2})_{oj}$ is the calculated output concentration of the calibration system used for the jth audit.

+An unsubscripted d represents the number of defects observed from n audits of a lot size of N.

Table 18. DATA FOR ANALYZER A

Date	NO_2 Level (ppm)	Analyzer Reading (Volts) Replicate		
		1	2	3
4/25/73	Zero NO_2	.003	.003	.005
	.10 NO_2	.188	.186	.187
4/27/73	Zero NO_2	.001	.001	.000
	.10 NO_2	.205	.204	.204
5/ 7/73	Zero NO_2	-.005	-.005	-.004
	.10	.190	.188	.186
	Zero NO_2	-.001	.000	.001
	.10	.185	.183	.185
	.40	.802	.803	.805
5/ 8/73	Zero NO_2	.002	.002	.001
	.10	.184	.184	.183
	.40	.784	.783	.785
5/ 9/73	Zero NO_2	.001	.001	.001
	.10	.185	.186	.185
	Zero NO_2	-.001	.000	.000
	.40	.786	.784	.785
5/11/73	Zero NO_2	.000	-.001	-.001
	.10	.185	.186	.184
	.40	.811	.812	.798
5/14/73	Zero NO_2	.002	.000	.001
	.10	.189	.189	.191
	.40	.798	.802	.799

Table 19. DATA FOR ANALYZER B

Date	NO_2 Level (ppm)	Analyzer Reading (Volts) Replicate 1	2	3
5/ 3/73	Zero NO_2	.003	-.014	.002
	.40 NO_2	.650	.643	.642
	Zero	-.015	-.011	.012
	.40	.629	.658	.630
	Zero	.007	.001	.006
	.40	.632	.643	.637
	Zero	.009	.003	.007
	.40	.640	.629	.640
	Zero	-.003	.012	.013
	.40	.644	.650	.675
	Zero	-.006	.004	[illegible]
	.40	.640	.671	.639
	Zero	-.003	-.002	-.002
	.40	.658	.676	.657
	Zero	.002	.011	.012
	.40	.667	.632	.655
	Zero	.004	-.010	.008
	.40	.678	.636	.679

Table 19 (CONTINUED)

Date	NO_2 Level (ppm)	Analyzer Reading (Volts) Replicate		
		1	2	3
5/ 4/73	Zero	.005	-.025	.005
	.10	.149	.152	.152
	.40	.799	.825	.816
*5/ 7/73	Zero	-.013	-.011	-.022
	.10	.136	.132	.147
	.40	.744	.760	.759
†5/ 8/73	Zero	.006	.012	.017
	.10	.135	.151	.157
	.40	.814	.811	.805
5/ 9/73	Zero	-.003	.014	.007
	.40	.642	.632	.634
5/11/73	Zero	.014	.007	.016
	.10	.158	.141	.138
	.40	.795	.791	.793
5/14/73	Zero	-.003	-.004	.007
	.40	.797	.756	.753

---

*at 115 Volts.

†at 125 Volts.

Table 20. ANALYSIS OF VARIANCE OF ANALYZER A DATA

Source of Variation (S.V.)	Sum of Squares (S.S.)	Degrees of Freedom (D.F.)	Mean Square (M.S.)
Zero NO_2 Data			
Total (Corrected for Mean)	127.6×10^{-6}	26	-
Between Days	117.6×10^{-6}	8	14.7×10^{-6}
Within Days	10×10^{-6}	18	0.6×10^{-6}
0.10 NO_2 Data			
Total (Corrected for Mean)	972×10^{-6}	23	-
Between Days	952.7×10^{-6}	7	136×10^{-6}
Within Days	19.3×10^{-6}	16	1.2×10^{-6}
0.40 NO_2 Data			
Total (Corrected for Mean)	1498×10^{-6}	14	-
Between Days	1359×10^{-6}	4	339.8×10^{-6}
Within Days	139×10^{-6}	10	13.9×10^{-6}

Table 21. ANALYSIS OF VARIANCE OF ANALYZER B DATA

Source of Variation (S.V.)	Sum of Squares (S.S.)	Degrees of Freedom (D.F.)	Mean Square (M.S.)
Zero NO_2 Data			
Total (Corrected for Mean)	2937×10^{-6}	20	-
Between Days	1761×10^{-6}	6	293.5×10^{-6}
Within Days	1176×10^{-6}	14	84×10^{-6}
0.10 NO_2 Data			
Total (Corrected for Mean)	876.7×10^{-6}	11	-
Between Days	258.7×10^{-6}	3	86.2×10^{-6}
Within Days	618.0×10^{-6}	8	77.3×10^{-6}
0.40 NO_2 Data			
Total (Corrected for Mean)	9741.7×10^{-6}	14	-
Between Days	7973.7×10^{-6}	4	1993.4×10^{-6}
Within Days	1768×10^{-6}	10	176.8×10^{-6}
0.40 NO_2 Data (All data taken on one day)			
Total (Corrected for Mean)	6757×10^{-6}	26	-
Between Replicates Within a Day	2767×10^{-6}	8	346×10^{-6}
Within Replicates	3990×10^{-6}	18	222×10^{-6}

Table 22. SUMMARY OF DATA ANALYSES - ANALYZER A

Volts

	ZERO NO_2	.10 NO_2	.40 NO_2
$\hat{\sigma}_W$	.00075 Volts	.0011	.0037
$\hat{\sigma}_B$	.0022	.0067	.0104
$\hat{\sigma}$	.0023	.0068	.011
$\overline{X}$	.0008	.188	.796
CV	-	3.6%	1.4%

Table 23. SUMMARY OF DATA ANALYSES - ANALYZER B

	ZERO NO_2	.10 NO_2	.40 NO_2	.40 NO_2 (Within one day)
$\hat{\sigma}_W$	.0092	.0088	.0133	.015
$\hat{\sigma}_B$	.0084	.0017	.0246	.0064
$\hat{\sigma}$	.012	.009	.028	.016
$\overline{X}$	-	.145	.788	.796
CV	-	6.2%	3.6%	2.0%

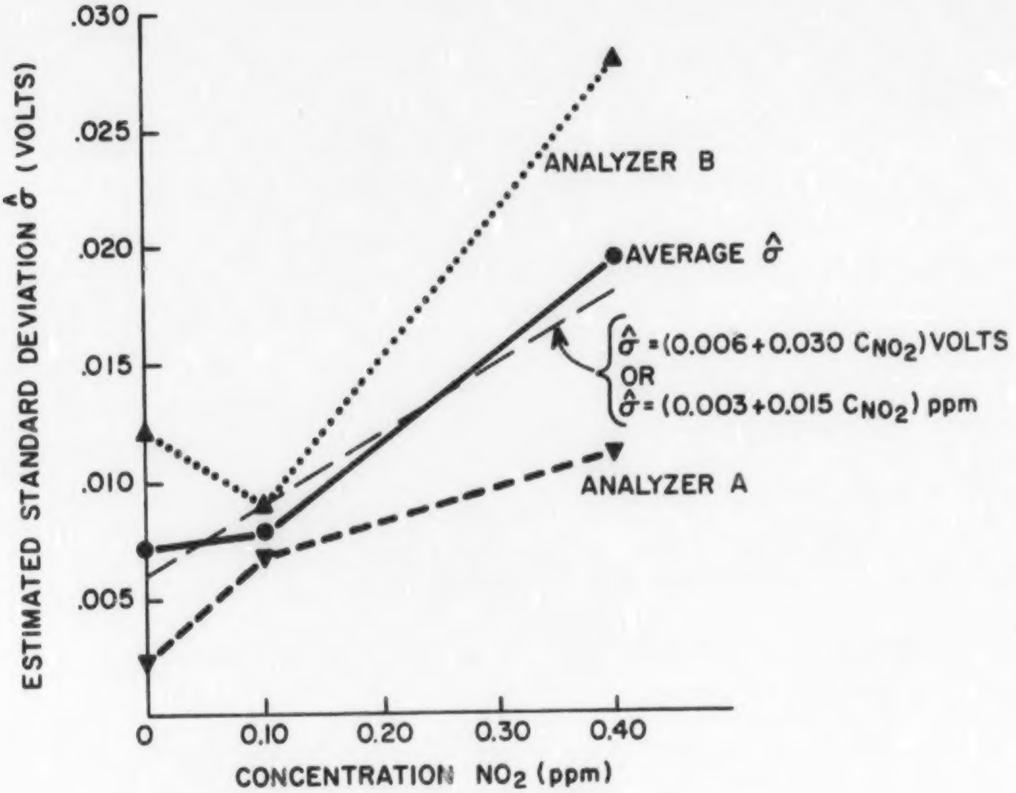


Figure 5: Estimated Standard Deviation of Analyzer Reading (volts) Versus NO_2 Concentration (ppm).

REFERENCE METHOD FOR THE CONTINUOUS MEASUREMENT OF CARBON MONOXIDE IN THE ATMOSPHERE (NON-DISPERSIVE INFRARED SPECTROMETRY)

General

Data quality of the reference method for the continuous measurement of carbon monoxide in the atmosphere as promulgated in the Federal Register can be estimated from three special checks. These checks are: 1) measurement of control samples, 2) water vapor interference checks, and 3) data processing checks (Ref. 20).

Arriving at Suggested Performance Standards

Suggested performance standards as given in the field document are reproduced in Table 24. Data used in arriving at performance standards and certain other control limits given in the field documents are discussed below.

The primary source of data on the precision and accuracy of the NDIR method for the continuous measurement of CO concentrations is given in Reference 21. A few remarks are given herein concerning the peculiarities of the test and their implications on the use of the data provided. Furthermore, the use of these data in the CO field document is described.

The collaborators were selected from experienced laboratories having a person with previous experience with the NDIR method of measuring the concentration of CO. Furthermore, each participant was given a preliminary test sample for analysis as a check of his capability; all "passed the test."

Test gases were provided to the test participants; these gases were analyzed both before and after the test. Good agreement between these analyses was indicated for all but 3 or 4 test gases. This indicates a need to check the test gases which are furnished by another laboratory prior to their extensive use in system calibration. Considerable precaution was taken during the test to obtain gases of sufficient accuracy and to safeguard these materials from contamination or deterioration in storage or use.

Test concentrations of CO were selected at 8, 30, and 53 mg/m^3. Effect of humidity was studied by analyzing the dry test gases after humidification. The test yielded 810 separate determinations--54 for each laboratory, 3 concentrations, 3 days, 3 determinations per day under both dry and humid conditions (3x3x3x2=54).

Of the 16 laboratories, 15 satisfactorily completed the tests. These laboratories constitute a sample of experienced rather than typical field laboratories in the onitoring network. Calibration curves were prepared independently on each of the three days. All arithmetic errors were corrected; few errors were noted as the method is relatively simple and not subject to such errors.

Statistical Analysis - Replication error is defined as the variation among successive determinations with the same operator and instrument on the same sample within time intervals short enough to avoid change of environmental factors and with no instrument manipulations other than zero adjustments.

Table 24: Suggested Performance Standards

Standards for Defining Defects

1. Measurement of Control Samples; $d_{1i} \geq \pm 2.2$ ppm
2. Water Vapor Interference Check; $d_{2i} \geq 1.7$ ppm

Standard for Correcting Data Processing Errors

3. Data Processing Check; $|d_{3i}| \geq 1$ ppm

Standards for Audit Rates

4. Suggested minimum auditing rates for data error; number of audits, $n = 7$; lot size, $N = 100$; allowable number of defects per lot, $d = 0$.
5. Suggested minimum auditing rates for data processing error; number of audits, $n = 2$; lot size, $N = 24$; allowable number of defects (i.e., $|d_{3i}| \geq 1$ ppm) per lot, $d = 0$.

Standards for Operation

6. If at any time $d = 1$ is observed (i.e., a defect is observed) for either d_{1i} or d_{2i}, increase the audit rate to $n = 20$, $N = 100$ until the cause has been determined and corrected.
7. If at any time $d = 2$ is observed (i.e., two defects are observed in the same auditing period), stop collecting data until the cause has been determined and corrected. When data collection resumes, use an auditing level of $n = 20$, $N = 100$ until no defects are observed in three successive audits.
8. If at any time either one of the two conditions listed below is observed, 1) increase the audit rate to $n = 20$, $N = 100$ for the remainder of the auditing period, 2) perform special checks to identify the trouble area, and 3) take necessary corrective action to reduce error levels. The two conditions are:
    a) two (2) d_{1i} values exceeding ± 1.4 ppm, or
       four* (4) d_{1i} values exceeding ± 0.7 ppm.
    b) two (2) d_{2i} values exceeding ± 1.0 ppm, or
       four* (4) d_{2i} values exceeding ± 0.5 ppm.

---

*A value of three was given in the field document; it should be changed to four.

The results indicate that this error is not affected by concentration or by humidity within the limits of the test, and thus all individual estimates are pooled to yield

$$\hat{\sigma}(\text{repl}) = 0.17 \text{ mg/m}^3 (0.148 \text{ ppm}) (286 \text{ degrees of freedom}).$$

All further analyses are made on averages of the three determinations made within one day on a given sample.

The use of a suitable drying agent or of refrigeration for maintaining constant humidity was satisfactory, whereas the use of narrow band optical filters alone was not satisfactory. In conclusion, humidity had no measurable effect on the accuracy of the method and did not appear to contribute significantly to the precision.

A linear model analysis was performed to estimate the several effects which contribute to the overall variation, precision, and accuracy of the reported results. The effects are assumed to be modeled by a linear additive effects model.

An analysis was performed to estimate the within day precision vs. the between- or among-day precision. These estimates are σ = σ (within day) = 0.17 mg/m^3, and σ_D = σ (among days) = 0.44 mg/m^3. All subsequent analyses were performed using the mean of three measurements on one sample on one day. The standard deviation of the average of three measurements within one day is σ (ave. of 3 measurements) =

$$\left[\sigma_D^2 + \frac{\sigma_W^2}{3}\right]^{1/2} = 0.45 \text{ mg/m}^3.$$

Straight lines were fitted to the measured concentrations vs. the average concentration for all laboratories. The means, slopes, and the standard errors of estimate (standard deviation of the observations about the fitted line) were determined for each laboratory. The means ranged from 29.73 to 32.07 mg/m^3; slopes, 0.928 to 1.0482; and the standard errors of estimate, 0.13 to 0.74 mg/m^3 (0.113 to 0.644 ppm). No outliers in these values were noted. The midpoint or average of the range of the standard error of estimate given above is 0.38 ppm. A conservative value of 0.33 ppm was recommended in the field document for checking calibration points. That is, at the 3σ level all calibration points should be within about ± 1.0 ppm (plus an additional allowance for error in the calibration gas) of the best-fit curve 99.7 percent of the time.

The next step in the analysis was to determine the correlation, if any, between the means and the slopes. There is a slight indication of correlation; however, even at the point where all laboratories tend to agree best, there is a significant variability in the results.

Finally, the components of variance for various sources of error are estimated.

$$\text{Variance (single measurement)} = \hat{\sigma}_D^2 + \hat{\sigma}_W^2 = 0.223$$

$$\hat{\sigma} \text{ (single measurement)} = 0.472 \text{ mg/m}^3.$$

Two errors are defined and estimated in the analysis using the method given in Ref. 22. Repeatability is defined as a quantity that is exceeded only about 5 percent of the time by the difference of two randomly selected test results obtained in the same laboratory. This is a different quantity than that used in the collaborative report on suspended particulate matter in that it is the product of 2.77 times the standard deviation. The factor 2.77 is the product of 1.96 (5 percent value of the standard normal deviate) and $\sqrt{2}$. (The difference of two observations has a variance of twice that of a single observation or a standard deviation $\sqrt{2}$ times that of a single value.) The reproducibility is a quantity exceeded about 5 percent of the time by the difference, larger less the smaller value, of two single results made on the same material (concentration) in two different randomly selected laboratories. These measures of repeatability and reproducibility are in some cases dependent on the concentration. In this case

$$\hat{\sigma}\text{ (repeatability)} = 0.55\text{ mg/m}^3$$

$$\hat{\sigma}\text{ (reproducibility)} = \left\{0.001\ x^2 - 0.039\ x + 1.1\right\}^{1/2}$$

where x is the concentration. For example, for $x = 8$ mg/m^3, $\hat{\sigma}$ (reproducibility) = 0.9 mg/m^3, or reproducibility = 2.77(0.9) = 2.5 mg/m^3. The results must be interpreted carefully because of the controlled conditions under which they were obtained and also because several outlying observations were eliminated in the analysis.

Interpretation of the Parameters - Interpretation of the results are:

(1). It is indicated that replication, taking more observations on the same day on the same sample, will in general be a waste of time. This is very often the case in practice; that is, the day-to-day variation often exceeds by a great deal the within-day sample variation, and hence taking additional samples on a day will in general not be cost beneficial. There are procedures for relating cost and variance; e.g., see Reference 23, which gives a procedure for estimating how many replicates, repeats (day-to-day), etc., should be taken to obtain an estimate of desired precision with minimum cost of measurements. In an auditing program in which a deviation is obtained between two measurements taken under conditions of replicates, the absolute deviation between two such measurements should be less than

$$\sqrt{2}\ (1.96)\ 0.17 = 0.47\text{ mg/m}^3,\quad 95\%\text{ of the time, or}$$

$$\sqrt{2}\ (2.58)\ 0.17 = 0.62\text{ mg/m}^3,\quad 99\%\text{ of the time.}$$

The values 1.96 and 2.58 assume that the estimate of the standard deviation 0.17 is known or based on an infinite number of degrees of freedom (D.F). If the D.F. is relatively small, say less than 25 or 30, then one should use the studentized range test, Reference 24.

(2). The day-to-day variation, involving changing environmental effects, as well as the short-time effects of instruments, as measured by the standard deviation is 0.57 mg/m^3. This measure of repeatability (on similar samples analyzed on different days) is useful in performing any audit check involving measurements made on different days but within the same laboratory. Thus if two samples are measured under these conditions, the measured values should differ by less than

$\sqrt{2}$ (1.96) 0.57 = 1.58 mg/m^3, 95% of the time, or

$\sqrt{2}$ (2.58) 0.57 = 2.08 mg/m^3, 99% of the time.

(3). The precision between laboratories is complicated by its dependence on the concentration level. This dependence was not evident in the previous measures. The comparison of measurements across laboratories such as that of a collaborative test used by an agency to check several labs involves not only the replication effect, and day-to-day effects, but also the variation between or among laboratories. The standard deviation is dependent on the concentration level ranging from 0.83 at 20 mg/m^3 to 1.53 at 60 mg/m^3 (and to 1.04 mg/m^3 at 0 mg/m^3). These values must be multiplied by the appropriate factors as given above for 95% and 99% limits of variation.

The suggested standard of 2.2 ppm for measuring control samples as given in Table 24 was arrived at as follows. The actual CO concentration of the sample is determined by the supplier and the measured value by the laboratory. Therefore, the standard deviation of the difference of two observations is applicable (see Appendix 4, page 144). The expected variability then is assumed to be greater than the repeatability of a single laboratory and less than or equal to the reproducibility value given above for among-laboratories since two independent measurements are involved. As a standard to be used when a quality assurance program is first started, the standard deviation for reproducibility at the point of highest precision, 20 mg/m^3, was suggested. The standard deviation of reproducibility at this CO concentration is 0.72 ppm (0.83 mg/m^3). Assuming that the variability in the known value of the control samples is much less than the variability of the measured value, the standard deviation of the difference was taken as the standard deviation of the measured value. The suggested performance standard for defining a defect when measuring control samples was 2.2 ppm (at the 3σ level). (All standards were based on a 0 to 50 ppm analyzer scale.)

The suggested standard for water vapor interference check was an estimate and not based on real data. There were two reasons for suggesting this value. They are:

(1) If a value much smaller than the normal measurement variability is used, there is an increased probability of declaring a defect due to water vapor interference when actually it is normal measurement error.

(2) If a larger value is allowed, it approaches the value used for measuring control samples and begins to significantly influence overall data variability.

The standard for correcting data processing errors was suggested as 1 ppm for a 50 ppm full scale range. This standard was suggested with dynamic strip chart traces in mind. If an intergrating flask is used to smooth out the CO spikes, a much lower standard should be used.

Accuracy - According to the results of Reference 21, a definite bias exists in the measurements. It also seems to depend on the concentration. On the average the measured results were 2.5% high. It is expected that this bias results from the use of calibration gases which exhibit significant variation with respect to their specified concentration. Thus it is concluded in Reference 21 that care must be taken in obtaining high quality calibration gases and protecting them from deterioration. On this basis it was recommended in the field document for CO that a check be made initially of a new calibration gas, and if the check indicates a possible discrepancy, further action should be taken to verify the stated concentration, e.g., a retest or a new calibration.

Recommendations

Water vapor control units which operate at a specific dew point should not be used in areas where the ambient dew point is frequently lower than that of the control unit.

The water vapor interference check as part of the auditing process is a must to ensure that the control units are properly used and maintained.

An integrating flask should be made a standard item of equipment on CO monitors. It is important in reducing data reduction errors when reading strip charts and is perhaps even more critical when an automatic digital data acquisition system is used.

SECTION III BACKGROUND INFORMATION APPLICABLE TO THE MANAGEMENT MANUALS OF FIELD DOCUMENTS

GENERAL

This section contains a series of brief descriptions of the procedures given in the management manual of the individual field documents. The organization of this section follows that of the last three field documents, namely, O_3, NO_2, and SO_2. In addition the figures in these manuals have been included in this final report for the convenience of the reader. An attempt is made here to give the necessary background information or reference to same for development of the tables and figures of the field documents. As appropriate, the assumptions and limitations concerning the statistical techniques are provided. In the case of the average cost computation, the sensitivity was determined of the results to the assumed costs and the percent of good quality data.

Section III is subdivided into three major subsections entitled SELECTING AUDITING SCHEMES, COST IMPLICATIONS, and DATA QUALITY ASSESSMENT. These are briefly described in the following paragraphs.

Two approaches to assessing or evaluating data quality were described in the field documents. One approach was to define a defective measurement in terms of some standard or permitted deviation from audited value. Once such a definition has been made, the problem becomes one of sampling, a selected number of measurements from a defined "lot" of measurements, e.g., 100 days, and making a decision based on the number of defective measurements in the sample. This approach was suggested because it would be simple to implement and would require very few computations and minimize the reporting of data. There is considerable flexibility in this audit approach. The curves and tables provided herein, and in the field documents, allow the manager to select the audit level and the number of permissible defects (0 or 1 is the extent of the tables, and is considered adequate for practical sampling levels). The audit level determines the confidence level concerning the percent of good measurements in the lot. The definition of a defective measurement should be consistent with the required data quality and the expected variation in the measurements, subject to the conditions given in the reference method. Background information pertaining to the selection of auditing schemes is given in the next section.

The cost data will be most difficult to estimate. However, the basic sampling approach is not very sensitive to the specific cost data as shown in the section entitled COST IMPLICATIONS. Thus it is not critical to estimate the costs in order to implement an auditing program. This cost analysis is given to provide a methodology for considering the overall or average cost of auditing the measurement process and then making a decision concerning the lot based on the number of defective measurements contained in the sample drawn from the lot.

An important approach for improving the data quality is to consider alternative strategies for improving the precision/accuracy of the measured concentrations. The alternative strategies may include changes in operating procedures, training programs; equipment changes or modifications; or improved environmental control. Each of the strategies can be evaluated as to the expected improvement in data quality and the expected additional cost. An example of this approach is provided in each field document. That is, four or five alternative strategies for improving the quality of the data are proposed, the expected improvement in data quality is hypothesized and the cost per 100 samples is estimated. The expected improvement in data quality as given in the field documents is typically based on judgment, since there was no opportunity to run experiments to estimate the improvements in the data quality. However, this opportunity exists for the monitoring agencies and it should be encouraged. Suggested improvements in procedures should be discussed and encouraged among the laboratory analysts, site operators, and supervisory personnel; and when judged to be a potential improvement, tested experimentally. The fact that an improved procedure costs dollars but results in improved quality data may ultimately reduce the average cost per item of valid data and save a great deal in costs relative to air quality control decisions to be based on these data.

The second approach to the assessment of data quality is one which estimates the variation in the results due to various aspects of the measurement process and then combines these results into one value or overall assessment. This approach is described in the section entitled DATA QUALITY ASSESSMENT and it is suggested in addition to the previous one, after experience has been gained in reporting the data on number of defectives and further knowledge has been gained concerning the more critical variables in the process. In this approach an attempt is made to assess the process by use of control samples and checking of data processing errors (reading charts, transcribing and computational errors), and making appropriate repeat determinations. By use of these data, the standard deviation of the measured results can be estimated by methods given in Appendix 4 entitled ESTIMATION OF THE MEAN AND VARIANCE OF VARIOUS COMPARISONS OF INTEREST. In the case of two of the pollutants, suspended particulate matter and SO_2, a model was formulated for relating the measured concentration to the several variables or factors of the measurement process. The models were developed starting with the deterministic equations for the concentrations as given in the reference methods. This modeling approach was taken because there was no simple, reliable, and cost efficient audit method to make an overall assessment. An example computation is given on pages 154 through 157 under the heading, "Means and Variance of Nonlinear Functions." Also see Appendices 1 and 2 for further details on these models and the subsequent analyses.

The use of control charts in improving the data quality is briefly described in Appendix 5. Appendix 5 does not attempt to give a complete discussion of the subject of quality control, but it does give appropriate references to some standard texts and indicates potential uses of control charts. The discussion also considers the consequences of "over correcting" the data. This occurs, for example when a check or calibration point is run daily and data are corrected daily in accordance with the measured value of the check. This results in a greater variation in reported results than if a more standard reference method were used in making data adjustments (Ref. 25).

SELECTING AUDITING SCHEMES

Three concepts pertinent to the first approach in the field documents are described in this section. The first is the computation of the probability of accepting a collection of measurements, referred to as a lot of N items, as valid given that there are D defective measurements in the lot, n items are sampled, and the number of defects permitted in the sample is c or less. The second concept is that of the level of confidence concerning the percentage of good measurements in the lot of N given that either 0 or 1 defect has been obtained in the sample of n measurements. Thirdly, the computation of the percent of defects in data reported as valid is computed as a function of the incoming data quality (prior to auditing) and the parameters of the auditing scheme.

Computation of the Probability of Accepting a Lot as a Function of the Sample Size and the Acceptable Number of Defectives

Auditing procedures, as recommended in the field documents, where certain check measurements are made and where the results of each check identified the air quality data as good or defective are analogous to what is commonly referred to as "sampling without replacement from a population in which each element of the population belongs to one of two classes". The hypergeometric probability distribution provides the probabilities of an item belonging to either class (in this case the two classes are good and defective) when drawn from a lot (population) of known size containing a given ratio of good to defective items. The parameters of a hypergeometric distribution are:

(1) the sample size n,

(2) the lot size (or population size) N, and

(3) the number of defective items in the lot D.

To illustrate how the hypergeometric distribution is used, suppose that a lot of N items contains D defective items (or measurements) and that a sample of n items is selected at random from the lot without replacing an item prior to drawing the next item. Assume that if c or less defectives are drawn, the lot is considered acceptable in quality, otherwise not acceptable. The probability of drawing a sample of size n with d defectives is given by

$$P\begin{pmatrix}\text{d defectives in}\\ \text{a sample of n}\end{pmatrix} = \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} ,$$

the hypergeometric distribution. The notation

$$\binom{D}{d} \text{ is defined as } \frac{D!}{d!\ (D-d)!} \ .$$

For example, if $n = 7$, $N = 100$, $D = 5$, $d = 0$, we have

$$\frac{\binom{5}{0}\binom{95}{7}}{\binom{100}{7}} = 0.6903.$$

If N is large with respect to n, say $N/n > 10$, then the binomial approximation to the above computation is adequate for most problems; i.e., if we let

$$\frac{D}{N} = p, \text{ and}$$

$$\frac{N-D}{N} = q \text{ say,}$$

we have

$$P\begin{pmatrix}\text{o defectives}\\ \text{in } n\end{pmatrix} \simeq \binom{n}{o}\ p^{o}\ q^{n} = \left(\frac{N-D}{N}\right)^{n} \ .$$

In the above example, this approximation gives

$$P\begin{pmatrix}\text{o defectives}\\ \text{in } 7\end{pmatrix} \simeq (0.95)^{7} = 0.70$$

which is very nearly equal to the above exact value. If the items are replaced in the lot before the next item is drawn (sampling with replacement), then the above formulation is exact; i.e., the binomial distribution applies.

If one permits more than 0 defectives, say c or less, for the lot to be acceptable, the probability of accepting the lot is the sum of the probabilities of c or less defectives i.e., if d is the observed number of defectives, we compute

$$P(d \leq c) = P(d = 0) + P(d = 1) + \ldots + P(d = c).$$

if c = 0, the computation in the example above for n = 7 becomes the probability of accepting the lot of items.

Table 25 below gives the results of the probability of d defectives for d=0, 1, 2, ...,6, n=7, N=100, D=5 and 15%. Figures 6 and 7 give similar results for varying sample size n(1 to 25), for N=50 and 100, respectively.

Table 25. P(d defectives)

	Data Quality	
d	D=5% Defectives	D=15% Defectives
0	0.6903	0.3083
1	0.2715	0.4098
2	0.0362	0.2152
3	0.0020	0.0576
4	0.00004	0.0084
5	≃ 0	0.0007
6	≃ 0	≃0

Confidence Interval Estimate for the Percentage of Good Measurements in a Lot

For this discussion, it is assumed that a sample of n measurements is selected at random from a lot of N measurements, and that the definition of a defective measurement has been made explicit, e.g., within ± 2 units of a prescribed value. It is desired to determine a confidence interval estimate of the percentage of good measurements in the lot (or vice versa, the percentage of defects in the lot). The standard statistical procedure employed for obtaining this confidence interval is similar to that for the binomial variable. (For example, see A. Hald, Ref. 26, pages 697-699). The procedure is to determine the percentage of defects D in the lot such that the probability of observing d or less defects in the sample is equal to the quantity, one minus the level of confidence. For example, suppose that 0 defects are observed in a sample of n=10 from a lot of N=10 measurements, and that an 80% confidence interval for the percentage of good (or defective) measurements is to be determined. If the percentage of defects in the lot is D, then the probability of 0 defects in the sample is

$$P(d=0) = \frac{\binom{D}{0}\binom{N-D}{n}}{\binom{N}{n}} .$$

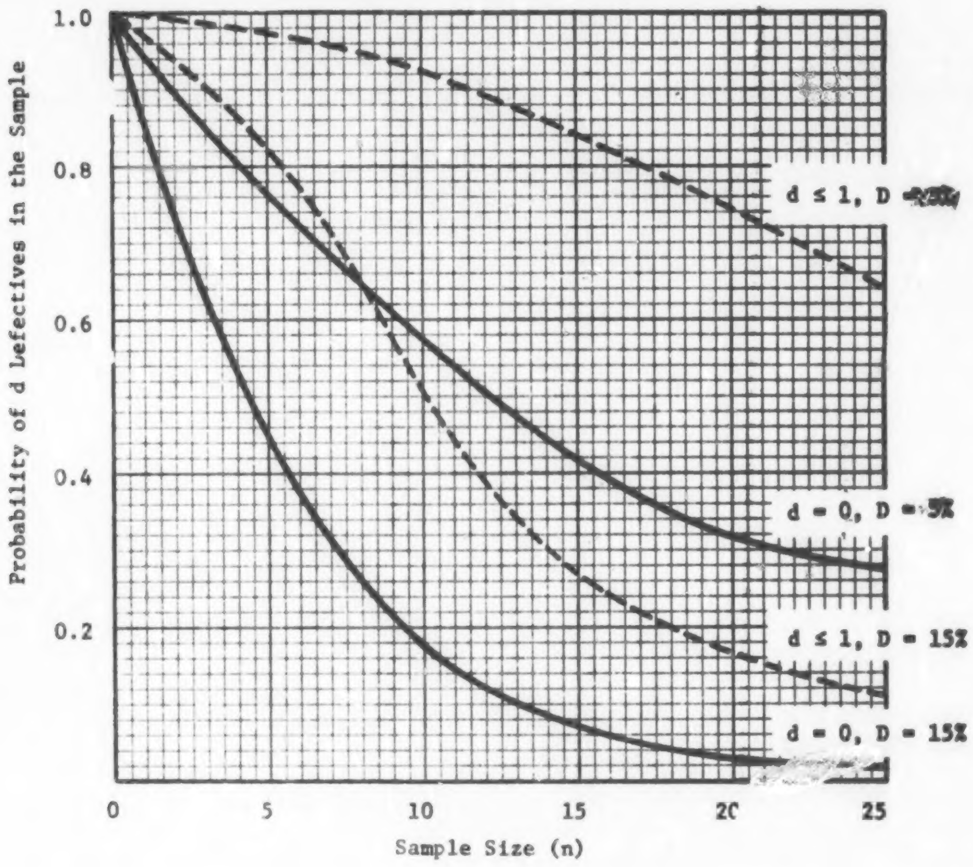


Figure 6: **Probability of d Defectives in the Sample If the Lot (N = 100) Contains D% Defectives**

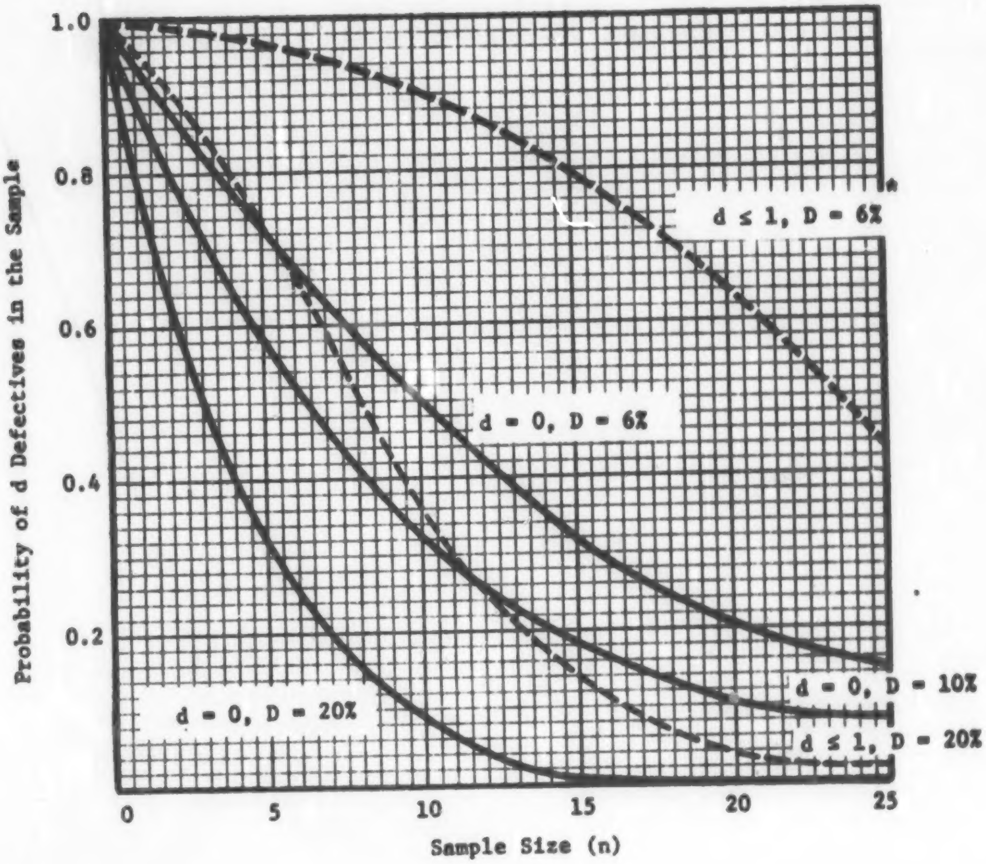


Figure 7: **Probability of d Defectives in the Sample If the Lot (N = 50) Contains D% Defectives.**

---

[*] **This graph is for a lot size of N = 50. Only whole numbers of defectives are physically possible; therefore, even values of D (i.e., 6, 10, and 20 percent) are given rather than the odd values of 5 and 15 percent as given in Figure 6.**

The value of D is then determined for which P(d=0) is equal to or smaller than 1 - 0.80 = 0.20. Because D is a discrete value, 0, 1, 2, ..., one increases D until the P(d=0) becomes less than 0.20. (In some cases, the value of D was chosen corresponding to P(d=0) being close to 0.20.) For example,

$$\text{If } D = 13,\ P(d=0) = \frac{\binom{13}{0}\binom{87}{10}}{\binom{100}{10}} = 0.231\ .$$

$$\text{If } D = 14,\ P(d=0) = \frac{\binom{14}{0}\binom{86}{10}}{\binom{100}{10}} = 0.204\ .$$

$$\text{If } D = 15,\ P(d=0) = \frac{\binom{15}{0}\binom{85}{10}}{\binom{100}{10}} = 0.181\ .$$

In this case D=14 or 14% defective measurements was taken as an upper 80% confidence limit on the percent of defects in the lot of N=100 measurements.

If one defective is observed; a similar approach is employed to determine the value of D using the sum of the probabilities of 0 and 1 defective, i.e

$$P(d=0) + P(d=1) = P(d \leq 1), \text{ or}$$

$$P(d \leq 1) = \frac{\binom{D}{0}\binom{N-D}{n}}{\binom{N}{n}} + \frac{\binom{D}{1}\binom{N-D}{n-1}}{\binom{N}{n}}$$

D is increased until the sum reaches the desired value, one minus the level of confidence. For example, if D=27, n=10, N=100

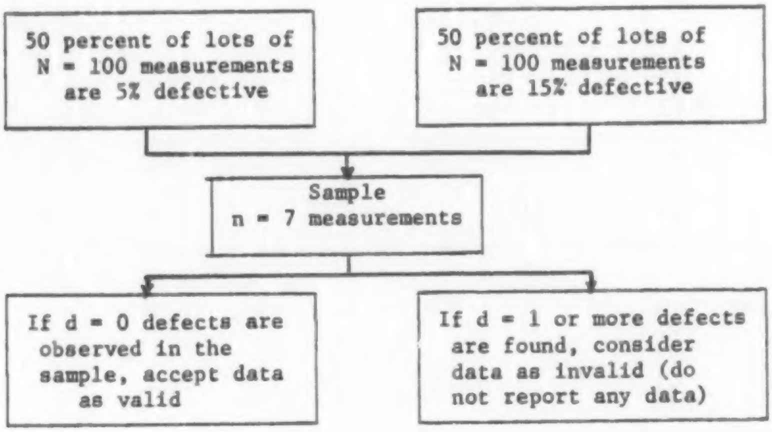
$$P(d=0) + P(d=1) = \frac{\binom{27}{0}\binom{73}{10}}{\binom{100}{10}} + \frac{\binom{27}{1}\binom{73}{9}}{\binom{100}{10}} = .0359 + 0.1515 = 0.184.$$

Thus, D=27 is taken as the upper 80% confidence limit for the percent defectives in a lot of N=100. Several computations of this type were made in order to generate the corresponding curves of percent of good measurements vs. n(sample size) and the specified confidence levels. Figure 8 and 9 give these results for 0 and 1 observed defective measurements, respectively.

Computation of the Percent Defective in the Reported Data

In order to estimate the percent of defects in reported data on a particular measurement, an assumption must be made about the quality of the data prior to the sampling and corrective actions taken following the sampling. Thus, if the distribution of percentage defects in the incoming lots of size N being sampled, the actions taken if either (1) c or less defects are found in the sample of n measurements, or (2) more than c defects are detected, are known, it is a straightforward but tedious calculation to determine the quality of the reported data.

Consider first a simple example of the incoming data quality being subdivided into two categories.


50 percent of lots of N = 100 measurements are 5% defective
50 percent of lots of N = 100 measurements are 15% defective
Sample n = 7 measurements
If d = 0 defects are observed in the sample, accept data as valid
If d = 1 or more defects are found, consider data as invalid (do not report any data)

The probability of accepting (rejecting) the lot of data (100 measurements) is given by the probability of 0 (1 or more) defectives; hence

$$P(0 \text{ defects} | 5\% \text{ defects}) = \frac{\binom{5}{0}\binom{95}{7}}{\binom{100}{7}} \approx 0.690$$

$$P(0 \text{ defects} | 15\% \text{ defects}) = \frac{\binom{15}{0}\binom{85}{7}}{\binom{100}{7}} \approx 0.308.$$

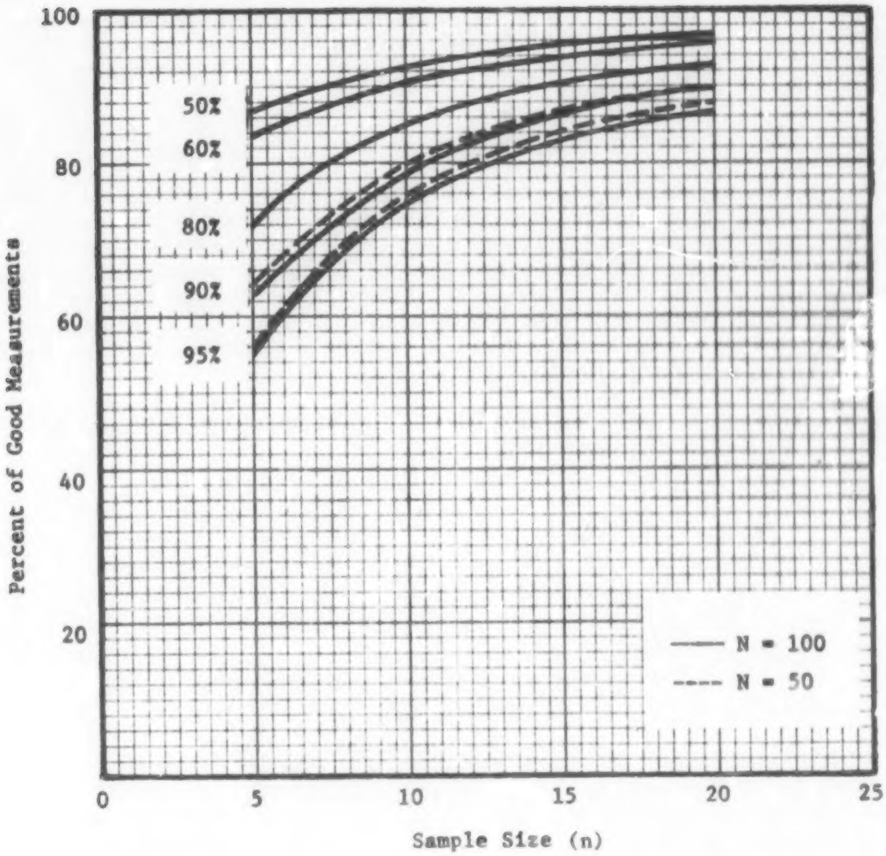


Figure 8: Percentage of Good Measurements Vs. Sample Size for No Defectives and Indicated Confidence Level

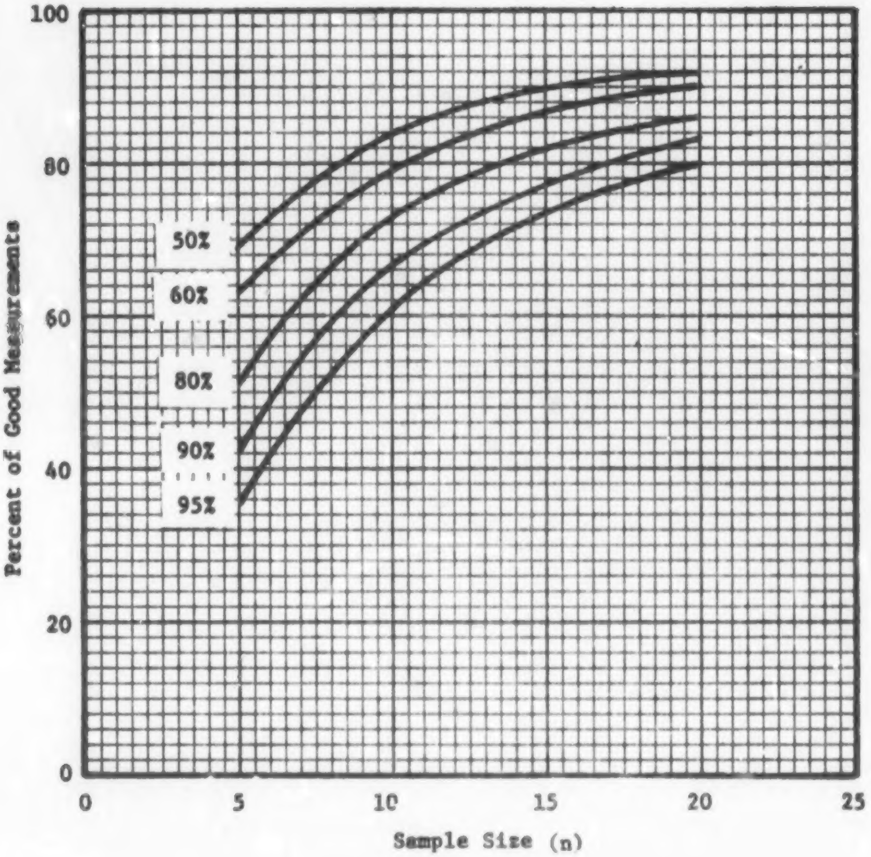


Figure 9: **Percentage of Good Measurements Vs. Sample Size for 1 Defective Observed and Indicated Confidence Level**

Lot Size = 100

Thus the average quality level, $\bar{p}$, of the reported data under the assumption that all lots declared invalid are eliminated is given by

$$\bar{p} = \frac{[0.5\ (0.05 \times P(0\ \text{def}\,|\,5\%\ \text{def}) + 0.5\ (0.15 \times P(0\ \text{def}\,|\,15\%\ \text{def})]}{[0.5\ P(0\ \text{def}\,|\,5\%\ \text{def}) + 0.5\ P(0\ \text{def}\,|\,15\%\ \text{def})]}$$

$$= \frac{[0.5\ (0.05 \times .69) + 0.5\ (0.15 \times 0.308)]}{0.5(.69 + 0.308)}$$

$$= \frac{0.0405}{0.499} = 0.081 \text{ or } 8.1\%.$$

In general if there are g% of lots of 100 which are good quality, q = (100 - g)% of lots which are poor quality, the above computation would be as follows.

$$\bar{p} = \frac{\left[\frac{g}{100}\left(0.05 \times P(0\ \text{def}\,|\,5\%\ \text{def})\right) + \left(\frac{q}{100}\right)\left(0.15 \times P(0\ \text{def}\,|\,15\%\ \text{def})\right)\right]}{\left\{\frac{g}{100}\left(P(0\ \text{def}\,|\,5\%\ \text{def})\right) + \left(\frac{q}{100}\right) P(0\ \text{def}\,|\,15\%\ \text{def})\right\}}$$

Several values of g are used below and the corresponding $\bar{p}$ determined for n = 7 and 17(c = 0 is assumed for both cases).

n = 7

g	$\bar{p}$(%)
90	5.5
80	6.0
70	6.6
60	7.3
50	8.1

n = 17

g	$\bar{p}$(%)
90	5.1
80	5.3
70	5.5
60	5.8
50	6.1

In assuming that all good lots contain 5% defects and all bad lots contain 15% defects, a rather restrictive assumption has been made. Now consider how the $\bar{p}$ is altered for a case in which the lot quality varies in accordance with the following distribution of percent defects.

	Good Lots(<10% defects)					Bad Lots (>10% defects)		
% defective	1	3	5	7	9	11	15	19
Percentage of Lots	10	10	10	10	10	15	20	15

With the assumed distribution, the average percent defects in the "good lots" is 5%, and for the "bad lots" is 15%.

	Good Lots (<10% defects)					Bad Lots (>10% defects)		
% defective	1	3	5	7	9	11	15	19
Prob(0 def in n=7)	0.92	0.81	0.69	0.60	0.52	0.45	0.308	0.24

Thus $\bar{p}$ = 9.1% vs. 8.1% obtained when 50% of the lots were 5% defective and 50%, 15% defective.

COST IMPLICATIONS

Pertinent cost considerations are described in this section for the purpose of illustrating methodologies. The conditional cost of accepting or rejecting a lot of data given a specified number of defects in the sample is considered in the following section. The concept of average cost is defined and the method given for computing this cost as a function of the quality of the data, costs of accepting poor quality data, rejecting good quality data, and the auditing cost. It is recognized that the assumed costs may not apply to a particular agency and that changes in the costs may affect the criteria used in the auditing scheme for accepting a lot of data as valid. Thus, a cost analysis is performed to determine the effect of changes in the costs as well as the assumed lot quality on the criterion for acceptance. No significant effect on the criterion was determined for the parameter changes employed in the analysis.

In the last subsection a methodology is given for performing cost tradeoffs, that is, a procedure for comparing alternative strategies for improving data quality (precision/accuracy) subject to estimated cost of implementing these strategies. When the costs and potential improvements in data quality are determined by a specific agency, the best strategy for improving the data quality may be selected using the methodology as suggested.

Bayesian Scheme for Computation of Costs

The auditing scheme can be translated into costs using the costs of auditing, rejecting good data, and accepting poor quality data. These costs may be very different in different geographic locations. Therefore, purely for purposes of illustrating a method, the cost of auditing is assumed to be directly proportional to the auditing level. For n = 7 it is assumed to be $155 per lot of 100. The cost of rejecting good quality data is assumed to be $600 for a lot of N = 100. The cost of reporting poor quality data is taken to be $800. To repeat, these costs given in Table 26 are assumed for the purpose of illustrating a methodology of relating auditing costs to data quality. Meaningful results can only be obtained by using correct local information.

Table 26. COSTS VS. DATA QUALITY

	Data Quality "Good" $D \leq 10\%$	 "Bad" $D > 10\%$
	Incorrect Decision	Correct Decision
Reject Lot of Data	Lose cost of performing audit plus cost of rejecting good quality data. (-$600 - $155)	Lose cost of performing audit, save cost of not permitting poor quality data to be reported. ($400 - $155)
	Correct Decision	Incorrect Decision
Accept Lot of Data	Lose cost of performing audit. (-$155)	Lose cost of performing audit plus cost of declaring poor quality data valid. (-$800 - $155)

Suppose that 50 percent of the lots have more than 10 percent defectives and 50 percent have less than 10 percent defectives. For simplicity of calculation, it is further assumed that the good lots have exactly 5 percent defectives and the poor quality lots have 15 percent defectives.

Suppose that n = 7 measurements out of a lot of N = 100 have been audited and none found to be defective. Furthermore, consider the two possible decisions of rejecting the lot and accepting the lot and the relative costs of each. These results are given in Tables 27 and 28.

Table 27. COSTS IF 0 DEFECTIVES ARE OBSERVED AND THE LOT IS REJECTED

		Correct Decision	Incorrect Decision	Net Value ($)
Reject Lot	D = 5%	---	$p_1 = 0.69$ $C_1 = -600 - 155$	$p_1C_1 = -\$521$
	D = 15%	$p_2 = 0.31$ $C_2 = 400 - 155$	---	$p_2C_2 = \$76$

Cost = $p_1C_1 + p_2C_2 = -\$445$

Table 28. COSTS IF 0 DEFECTIVES ARE OBSERVED AND THE LOT IS ACCEPTED

		Correct Decision	Incorrect Decision	Net Value ($)
Accept Lot	D = 5%	$p_1 = 0.69$ $C_3 = -155$	--- ---	$p_1C_3 = -\$107$
	D = 15%	---	$p_2 = 0.31$ $C_4 = -800 - 155$	$p_2C_4 = -\$296$

Cost = $p_1C_3 + p_2C_4 = -\$403$

The value p_1 (p_2) in the above table is the probability that the lot is 5% (15%) defective given that 0 defectives have been observed. For example,

$$p_1 = \frac{\begin{pmatrix}\text{probability that the lot is 5\% defective}\\ \underline{\text{and}}\ \text{0 defectives are observed}\end{pmatrix}}{p\begin{pmatrix}\text{lot is 5\% defective }\underline{\text{and}}\\ \text{0 defectives observed}\end{pmatrix} + p\begin{pmatrix}\text{lot is 15\% defective }\underline{\text{and}}\\ \text{0 defectives observed}\end{pmatrix}}$$

$$= \frac{0.5(0.69)}{0.5(0.69) + 0.5(0.31)} = 0.69 .$$

$$p_2 = \frac{\begin{pmatrix}\text{probability that the lot is 15\% defective}\\ \underline{\text{and}}\ \text{0 defectives are observed}\end{pmatrix}}{p\begin{pmatrix}\text{lot is 5\% defective }\underline{\text{and}}\\ \text{0 defectives observed}\end{pmatrix} + p\begin{pmatrix}\text{lot is 15\% defective }\underline{\text{and}}\\ \text{0 defectives observed}\end{pmatrix}}$$

$$= \frac{0.5(0.31)}{0.5(0.31) + 0.5(0.69)} = 0.31 .$$

It was assumed that the probability that the lot is 5% defective is 0.5. The probability of observing zero defectives, given the lot quality is 5% or 15%, can be read from the graph of Figure 6.

A similar table can be constructed for 1, 2, ..., defectives and the net costs determined. The net costs are tabulated in Table 29 for 1, 2, 3, and 4 defectives. The resulting costs indicate that the decision preferred from a purely monetary viewpoint is to accept the lot if 0 defectives are observed and to reject it otherwise. The decision cannot be made on this basis alone. The details of the audit scheme also affect the confidence which can be placed in the data qualification; consideration must be given to that aspect as well as to cost.

Table 29. COSTS IN DOLLARS

Decision	d = number of defectives				
	0	1	2	3	4
Reject Lot	-445	-155	+101	+207	+244
Accept Lot	-403	-635	-839	-928	-952

Computation of Average Cost

The average cost associated with a given audit procedure, associated decision criterion, and assumed costs for each decision is obtained by multiplying the probability of each possible decision by the associated cost and summing over all the possible cases. The general computation procedure is given by the formula.

Average Cost = P (lot is good quality and is accepted) Cost (Audit)

+ P (lot is good quality and is rejected) Cost (rejecting good data and cost of audit)

+ P (lot is poor quality and is accepted) Cost (accepting poor quality data and cost of audit)

+ P (lot is poor quality and is rejected) Cost (Savings of rejecting poor quality data less cost of audit)

Each of the probabilities in the above is computed by a formula similar to the following.

P (lot is good quality and is accepted) = P (lot is good quality) × P (lot is accepted|lot is good quality)

The notation $P(A|B)$ is read "probability that A occurs given that B has occurred; e.g., in the last expression above, the probability that the lot is accepted given that it is of good quality, say 5% defective measurements in the lot. Each probability is the product of the probability that the lot quality is as specified by the conditional probability that the lot is accepted (or rejected) given the specified lot quality. The four probabilities must add up to one (1) since all the possible situations are enumerated.

Table 30 below contains the detailed computations for an example in which the following assumptions are made.

Data Quality

Poor Quality Data: $D = 15\%$ defects.

Good Quality Data: $D = 5\%$ defects.

50% of the Lots are Good and 50% Poor Quality Data.

Decision Criterion

Sample 7 measurements from a lot of $N = 100$ measurements.

Accept lot if $d = 0$ defectives are observed.

Associated Costs

The costs are given in Table 26, page 86.

Table 30: Overall Average Costs for One Acceptance-Rejection Scheme

Decision	Good Lots D = 5%	Bad Lots D = 15%	Average Costs
Reject any lot of data if 1 or more defects are found.	q_1=0.5(0.31)=0.155 C_1 = -$755	q_2=0.5(0.69)=0.345 C_2 = $245	$q_1C_1+q_2C_2$ = -$32
Accept any lot of data if 0 defects are found.	q_3=0.5(0.69)=0.345 C_3 = -$155	q_4=0.5(0.31)=0.155 C_4 = -$955	$q_3C_3+q_4C_4$ = -$202

Average Cost = -$234

The cost -$234 is the (weighted) average cost of the four possible situations. The weights are the corresponding probabilities of each occurrence.

The average cost -$234 can also be obtained from the results of Table 29, page 88, multiplying the cost associated with each decision for the observed number of defectives d by the probability that d occurs and then summing the products over all the possible cases. This computation is given below in Table 31.

Table 31: Overall Average Costs

No. of Defectives	Decision Rule	Costs ($) from Table 3	Prob(d)	Cost × Prob(d)
d = 0	Accept	- 403	0.50	-$201.5
1	Reject	- 155	0.34	- 52.7
2	Reject	101	0.1255	12.6
3	Reject	207	0.030	6.2
4	Reject	244	0.0042	1.0
		Totals	0.9997	-$234.4

Using the method of Table 30, the average cost was computed for sample sizes n = 0 through n = 20. These results and other relevant information is summarized in Figure 10.

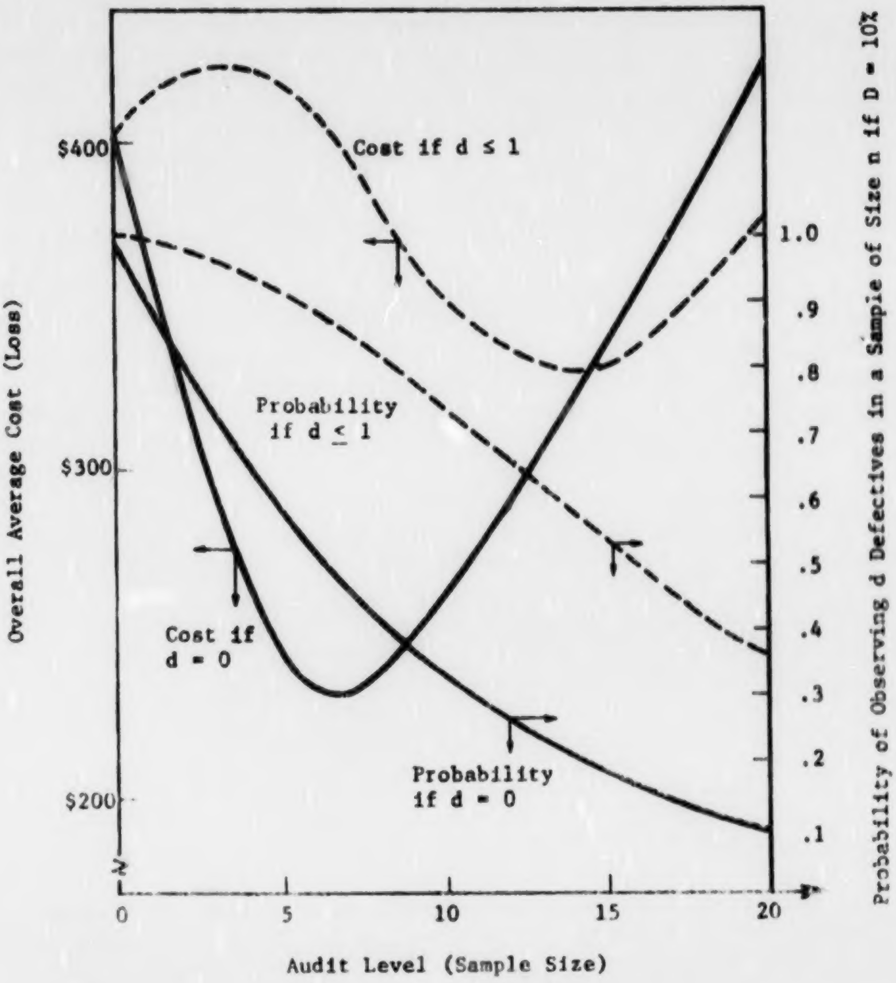


Figure 10: Average Cost Vs. Audit Level
(Lot Size N = 100)

Cost Analysis

In this subsection the effect of varying certain assumptions (parameters) used to obtain the average cost are considered. Recall that from the previous subsection the average cost is given as a function of (1) the percent defects in the good and poor quality lots (5 and 15% were assumed), (2) the percent of lots which are of good quality (50% was assumed), (3) the individual costs (audit cost, cost of rejecting good data, cost of accepting poor quality data, etc.), and (4) the variables associated with the sampling (lot size, sample size, and number of permissible defects in the sample). To vary all of the parameters simultaneously would require extensive computation. Hence one parameter at a time will be varied just to indicate the sensitivity of the results to each assumption.

Case 1: Vary the Percentage of Lots of Good Quality

Consider first the percentage of the lots which are of good quality. Let this percentage vary from the 50% assumed in the analysis in the field documents to 90%, using one intermediate value, 70%. All other parameters will be held constant. The effect of this change on the average cost is given in the following table.

Table 32: Quality of Incoming Data and Associated Average Cost

Percentage of Lots Which are of Good Quality	Average Cost
50	-$234
70	-$277
90	-$320

The above costs assume that the acceptance rule is to sample n = 7 measurements from N = 100 and to accept all data in the lot if no defects are observed; otherwise, reject the lot as unsatisfactory data quality. Now examine the costs associated with the Bayesian scheme for acceptance, conditional on the number of defects, following the procedure described in the previous subsection starting on page 86. In the notation of the text, two probabilitias are computed, p_1 (and p_2), for the conditional probabilities that the sample is from a good lot (poor quality lot) given that d defects have been observed in the sample. Assuming 70% of the lots are good, then

$$p_1 = \frac{0.7(0.69)}{0.7(0.69) + 0.3(0.31)} = 0.84$$

$$p_2 = \frac{0.3(0.31)}{0.3(0.31) + 0.7(0.69)} = 0.16 \quad .$$

If 0 defects are observed, and if the decision is to reject the lot, then the conditional cost is

$$\text{Cost} = p_1C_1 + p_2C_2 = 0.84(-755) + 0.16(245) = -\$595.$$

If the decision is to accept the lot, the cost is

$$\text{Cost} = p_1C_3 + p_2C_4 = 0.84(-155) + 0.16(-955) = -\$283.$$

On the basis of these costs, one would prefer to accept the lot if zero defects are observed (the same as in the field document).

If 1 defect is observed, the probabilities p_1 and p_2 become 0.61 and 0.39, respectively, and the corresponding costs associated with rejection and acceptance become -\$365 and -\$467, respectively, indicating that it is cost effective to reject the lot. Continuing the analysis in this manner for d = 2,3 and 4 defects we summarize the results in the following table for d = 0, 1, 2, 3, and 4.

Table 33: Summary Information

	d = number of defects observed in sample n = 7 from a lot of N = 100					
	0	1	2	3	4	
Reject the Lot	-\$595	-\$365	-\$35	+\$170	≈+\$245	
Accept the Lot	-\$283	-\$467	-\$731	-\$895	≈-\$955	
p_1	0.84	0.61	0.28	0.075	≈0	
p_2	0.16	0.39	0.72	0.925	≈1	
						Total
P (d defects)	0.576	0.312	0.090	0.019	0.003	1.000
Cost*	-\$283	-\$365	-\$35	+\$170	+\$245	
Cost x P(d defects)						-\$276

---

*Cost is the value corresponding to the preferred decision rule, i.e., accept for d=0, reject otherwise.

The above table gives the costs conditional on rejecting/accepting the lot, the values of p_1 and p_2, the probability that d defects will be observed, and the computation of the average cost.

In summary the change from 50% good quality lots to 70% good quality lots has not altered the general conclusions; only the relative costs have changed. Similar conclusions hold for 90% good quality lots. It should be noted that the average cost is increasing.

Case 2: Vary the Cost of Accepting Poor Quality Data

The cost of accepting poor quality data was taken to be -\$800 in the field documents. Let us vary this cost to take values of -\$400 and -\$600. The cost of rejecting the lot given d defects will not change from that given in the field document. However, this cost is repeated here with that for the three costs of accepting the lot.

Table 34: Effect of varying the Cost of Accepting Poor Quality Data

	d = number of defects in sample		
Decision	d=0	d=1	d=2
Reject the Lot	-\$445	-155	+101
Accept the lot (C_4 = -\$800)	-403	-635	-839
Accept the lot (C_4 = -\$600)	-341	-515	-669
Accept the lot (C_4 = -\$400)	-279	-395	-497

In summary the effect of varying the cost of accepting poor quality data from -\$400 to -\$800 has not altered the preferred decision rule, that is, to accept the lot if 0 defects in the sample and reject otherwise.

Case 3: Vary the Cost of Rejecting Good Data

In the field document the cost of rejecting good quality lots was taken to be -\$600, in this study we vary the cost from -\$400 to -\$800, similar to the manner above. In this case the costs of accepting the lot for various observed number of defects is not changed from that obtained in the field documents. The audit cost is -\$155 and is added to the cost of rejecting good quality lots to obtain C_1, the cost used in Table 35.

Table 35: Effect of Varying the Cost of Rejecting Good Quality Data

Decision Rule	d = number of defects in sample	
	d=0	d=1
Reject the lot (C_1 = -$555)	-$307	-$75
Reject the lot (C_1 = -$755)	-$445	-$155
Reject the lot (C_1 = -$955)	-$583	-$235
Accept the Lot	-$403	-$635

Case 4: Vary the Percent Defects in the Poor Quality Lots

It was assumed in the field documents that all poor quality lots contained 15 percent defects and the good quality lots contained 5 percent defects. Now let the percent defects in the poor quality lots take values of 20 and 30 percent, respectively. The results of these computations are summarized in the following table.

Table 36: Effect of Varying the Percent Defects in Poor Quality Lots

	d = number of defects in sample					
	0	1	2	3	4	5
Good Quality Lot						
P(d|D=5)	0.69	0.27	0.036	0.002	≃0	≃0
Poor Quality Lot						
P(d|D=15)	0.31	0.410	0.215	0.058	0.008	≈0
P(d|D=20)	0.20	0.375	0.285	0.113	0.025	≈0
P(d|D=30)	0.075	0.246	0.330	0.233	0.094	.021
P_1	0.775	0.418	0.113	0.017	-	-
P_2	0.225	0.582	0.887	0.983	-	-
Cost(reject lot)	-$530	[-$175]	[+$132]	-	-	-
Cost(accept lot)	[-$335*]	-$619	-$865	-	-	-

---

*Preferred decision rule implied by lesser costs denoted by brackets.

In summary the decision rule was altered in only one instance; that is, when the cost of rejecting good quality data was lowered to -$400, the cost of rejecting the lot for d = 0 was less than that of accepting the lot. Hence the preference to reject the lot even if d = 0. However, this would be an unacceptable decision rule for there would be no need to take any data. In practice this particular cost, i.e., for rejecting all data, should be large particularly because there is no evidence of poor quality data, i.e., d = 0 defects were observed.

Case 5: Vary the Audit Level or Sample Size

The variation of cost with sample size while holding fixed the following parameters: N, percent of good quality lots, and costs of sampling, rejecting good quality data, accepting bad quality data, etc., is given in Figure 10, page 91.

Cost Trade-Off Procedures

This section illustrates a methodology for comparing alternative strategies for improving the precision/accuracy of data. In practice it is necessary to experiment with the alternative strategies in order to obtain specific estimates of how much improvement, if any, can be expected and to relate the precision to the cost.

In order to illustrate the methodology, the strategies proposed for the measurement of the concentration of SO_2 are repeated here for convenience.

The assumed values of the standard deviations and biases for each strategy and audit are not based on actual data, except where such data are given for the reference method. Four alternative strategies (A1, A2, A3 and A4) and various combinations of these actions are considered. The added cost (above that of the reference method A0) for implementing each strategy is estimated on the basis of each set of 100 24-hour samples. The assumed data are given in Table 37 and the plot of added cost versus the estimated mean square error (expressed in percent) is given in Figure 11. Procedures for combining the biases and standard deviations to obtain an overall assessment are given in the next section.

Suppose that it is desired to make a statement that the true SO_2 concentration is within 12% of the measured concentration (for simplicity of discussion all calculations are made at a true concentration of 380 $\mu g/m^3$) with approximately 95 percent confidence. Minimal cost control equipment and checking procedures are to be employed to attain this desired precision. Examining the graph in Figure 11 of cost versus precision, one observes that A2 is the least costly strategy that meets the required goal of $2MSE \leq 12$ or $MSE \leq 6$ percent. The mean square error (MSE $= \sqrt{\sigma^2 + \tau^2}$ is used in this analysis as a means of combining the bias (τ) and standard deviation (σ) to obtain a single measure of the overall dispersion of the data. The assumed values of the MSE's of the measured concentrations of SO_2 for the alternative courses of action are given in Table 37. The costs for the various alternatives are given as the ordinate axis in Figure 11.

Suppose that it is desired that MSE be less than 4% and that the cost of reporting poor quality data increases rapidly for MSE greater than 4%, then strategy A6 = (A2 + A4) appears best because it meets the goal of MSE being less than 4% its costs of implementation is \$400/100 samples. However, strategy A5 costs \$215 to implement and results in a cost of about \$40 for reporting poor quality data; an overall cost of \$255 compared to \$400 for A6. Based on the assumed values A5 would be best. This example demonstrates the need for the manager to obtain estimates of the improvements in data quality which can be attained through various actions. This assumption is illustrated by the cost curve given by the solid line in Figure 11. For any alternative strategy, the cost of reporting poor quality data is given by the ordinate of this curve (the solid curve) corresponding to the strategy.

Table 37. ASSUMED STANDARD DEVIATIONS FOR ALTERNATIVE STRATEGIES[1]

		A0	A1	A2	A3	A4	A5[2]	A6[2]	A7[2]
1. Flow rate	$\bar{d}_1$	−4	−4	−4	−4	−3	−3	−1	−1
	σ_1	σ_1	$0.6\sigma_1$	σ_1	σ_1	$0.7\sigma_1$	$0.42\sigma_1$	$0.7\sigma_1$	$0.42\sigma_1$
2. Control sample	$\bar{d}_2$	0	0	0	0	0	0	0	0
	σ_2	σ_2	σ_2	$0.7\sigma_2$	$0.8\sigma_2$	$0.8\sigma_2$	$0.8\sigma_2$	$0.56\sigma_2$	$0.56\sigma_2$
3. Data processing	$\bar{d}_3$	0	0	0	0	0	0	0	0
	σ_3	σ_3	σ_3	σ_3	σ_3	$0.7\sigma_3$	$0.7\sigma_3$	$0.7\sigma_3$	$0.7\sigma_3$
σ_T (%)		5.0	4.7	5.1	4.7	3.8	3.5	3.8	3.5
Negative bias = −(%)		4	4	1	4	3	3	1	1
MSE(%)[4]		6.5	6.2	5.2	6.2	4.8	4.6	3.9	3.7
Added cost ($) per 100 samples		0	15	200	260	200	215	400	415

[1] $\sigma_2 = 0.4\ \mu g\ SO_2$, for a 24-hour sample at 380 $\mu g\ SO_2/m^3$ where 0.32 m^3 of air is sampled, $\sigma_2 = 0.4\ \mu g\ SO_2$ is equivalent to a standard deviation of 3.3%.

σ_1 and σ_3 are assumed to be 2.5% and 3% respectively of the average or mean value, $\bar{X} = 380\ \mu g\ SO_2/m^3$.

$\bar{d}_1$ is also expressed as the bias in % of the mean concentration, $\bar{X} = 380\ \mu g\ SO_2/m^3$.

[2] A5 = A1 + A4, A6 = A2 + A4, A7 = A1 + A2 + A4.

[3] $\sigma_T = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}$ = percent variation in measured concentration of SO_2, calculated at a mean concentration of 380 $\mu g/m^3$.

[4] MSE (%) $= \sqrt{\sigma_T^2 + \tau^2}$.

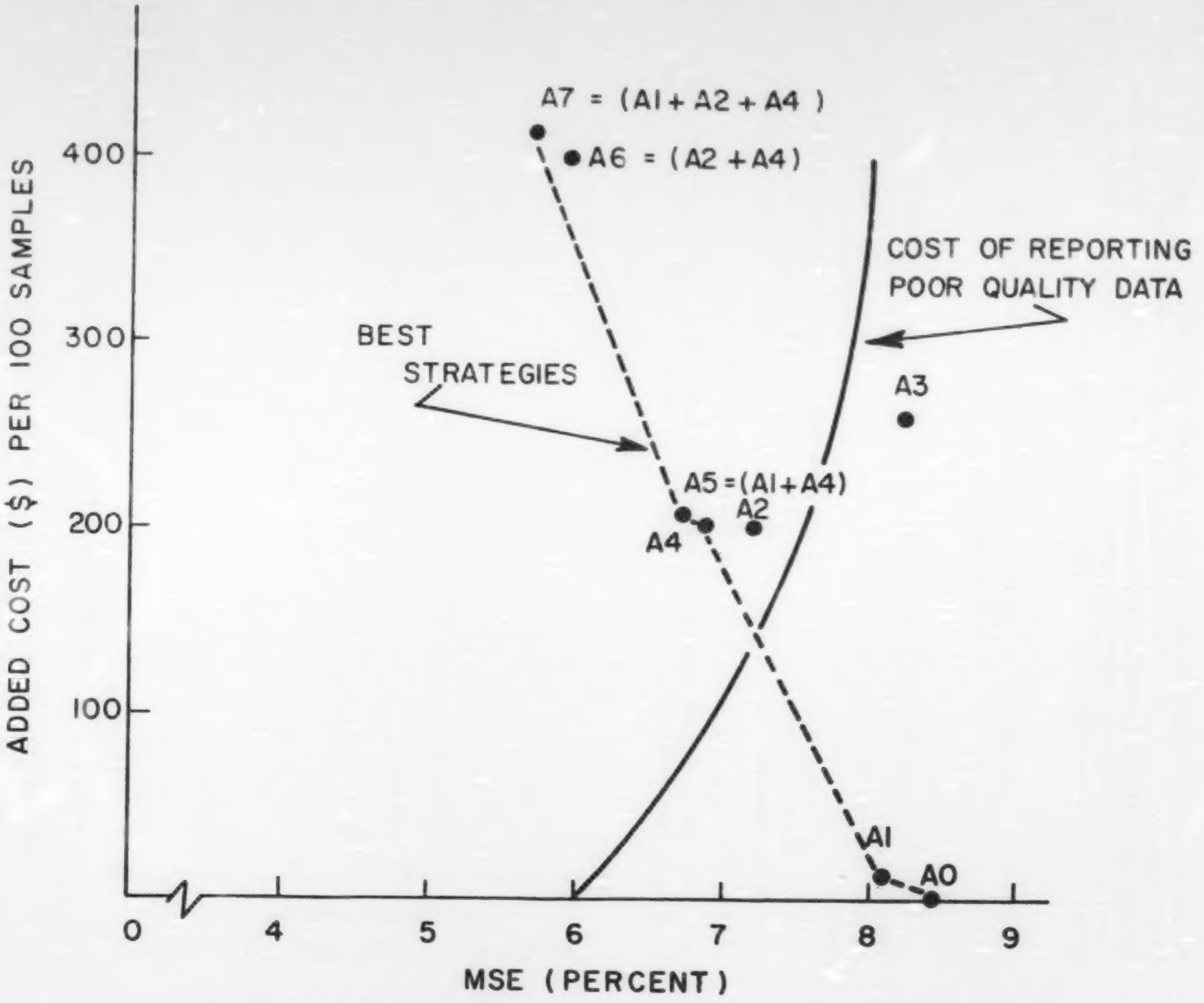


Figure 11: Added Cost ($) vs. MSE (%) for Alternative Strategies

To illustrate a technique which aided in the development of Table 37 and Figure 11, consider the following linear approximation to the model as derived from the sensitivity analysis (see Appendix 2 for a detailed discussion) which includes the most important variables and explains almost all of the variation in the measured concentration, C, of SO_2:

$$C = 378.8 + 12{,}630X(1) + 1156X(3) - 1156X(4) + 37.9X(5) + 394.6X(6) - 4.795X(13) + \dots .$$

To improve the precision/accuracy of the measured concentrations requires implementation of data quality control procedures applicable to any one of the important variables or to a combination of them. For example, the latter would result from implementation of action A7 (Figure 11) which is the sum of 3 actions and improves the precision of more than one variable. Assume that the cost of implementing a control procedure for any one variable X(j) is C_j, and for two or more variables, C_{jk}, e.g., if X(j) and X(k) were being controlled. In practice there may be several levels of control which are desirable to consider, and a function such as that shown below in Figure 12 is applicable.

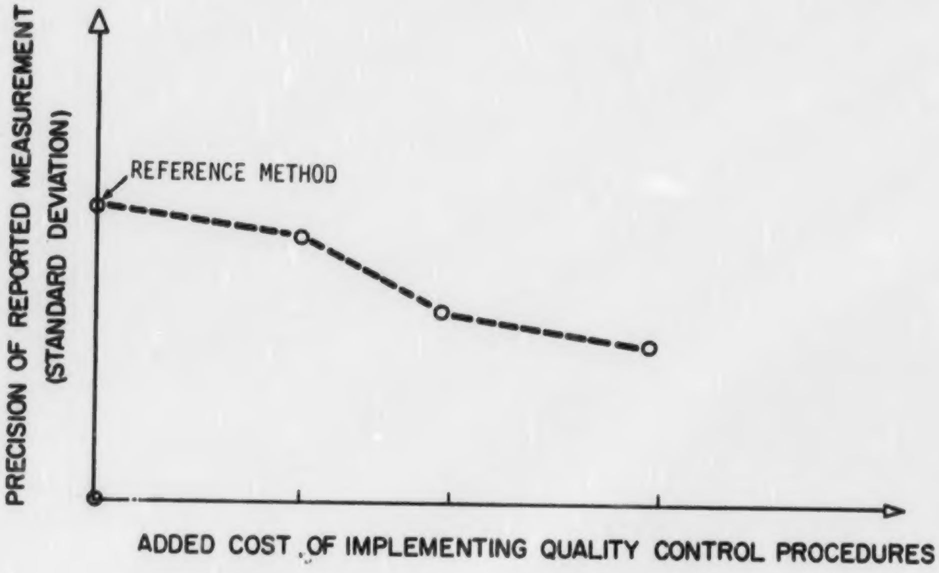


Figure 12: Precision of Reported Measurements Vs. Cost of Quality Control Procedure

It is desired to describe a methodology to select which quality control procedures to implement for a given level of resources or which procedures to implement to minimize cost while achieving a fixed degree of overall precision/accuracy. For simplicity of discussion, assume that there is one level of implementation for each variable, with associated cost C_j for variable X(j); i.e., one value other than the reference value for the function in Figure 12.

With a small number of variables and a small number of control strategies, it is easy to enumerate all cases, including combinations, as desired and obtain a relationship of overall precision of C versus the added cost C_T. The overall precision is given by the variance estimate

$$s^2(C) = 12.630^2\ s^2\{X(1)\} + 1156^2(s^2\{X(3)\} + s^2\{X(4)\}$$

$$+ 37.9^2\ s^2\{X(5)\} + 394.6^2\ s^2\{X(6)\} + 4.795^2\ s^2\{X(13)\} + \ldots.$$

or the standard deviation of the estimated C, s(C)

$$s(C) = \sqrt{s^2(C)}$$

where s {X(j)} is the precision (standard deviation) of the measurement of X(j). Figure 13 gives the form of the relationship.

To use the result of Figure 13 to determine the total cost corresponding to a desired level of precision $s_0(C)$, one uses the procedure corresponding to cost C_0. (See solid lines.) On the other hand if one has resources C_0^*, the best precision that can be obtained is $s_0^*(C)$. (See dashed lines.) Although the above structure and example are very simple, the approach illustrated is applied to a more complex problem. In general terms the problem might be formulated as follows.

Problem 1

minimize $s^2(C)$ or s(C); subject to given total added cost $C_T \leq A$

or

Problem 2

minimize C_T; subject to $s(C) \leq S$.

The mean square error may be substituted for s(C) in the above problem statements.

If the relationship of C_j to $s(X(j))$ is known for each variable or combinations thereof, then one needs a cost trade-off procedure, better known in the literature as a mathematical programming procedure, for solving such a problem. The cost is a linear function (additive) of the corresponding cost elements, and the precision $s(C)$ is a nonlinear function but a very simple one. A computer program can easily be structured to solve the problem for all cases of interest if the possibilities are discrete in number, or a LaGrangian multiplier approach can be used if a function form is given relating C_j to $s(X(j))$. More complex problems can be solved using heuristic search techniques and dynamic programming methods. Nonlinear programming techniques are also available for complex problems and functions. However, it is felt that a reasonably simple computer program which enumerates the function of $s(C)$ versus C_T would be adequate for problems 1 and 2 as stated above considering the state-of-the-art in estimating the costs and associated precisions.

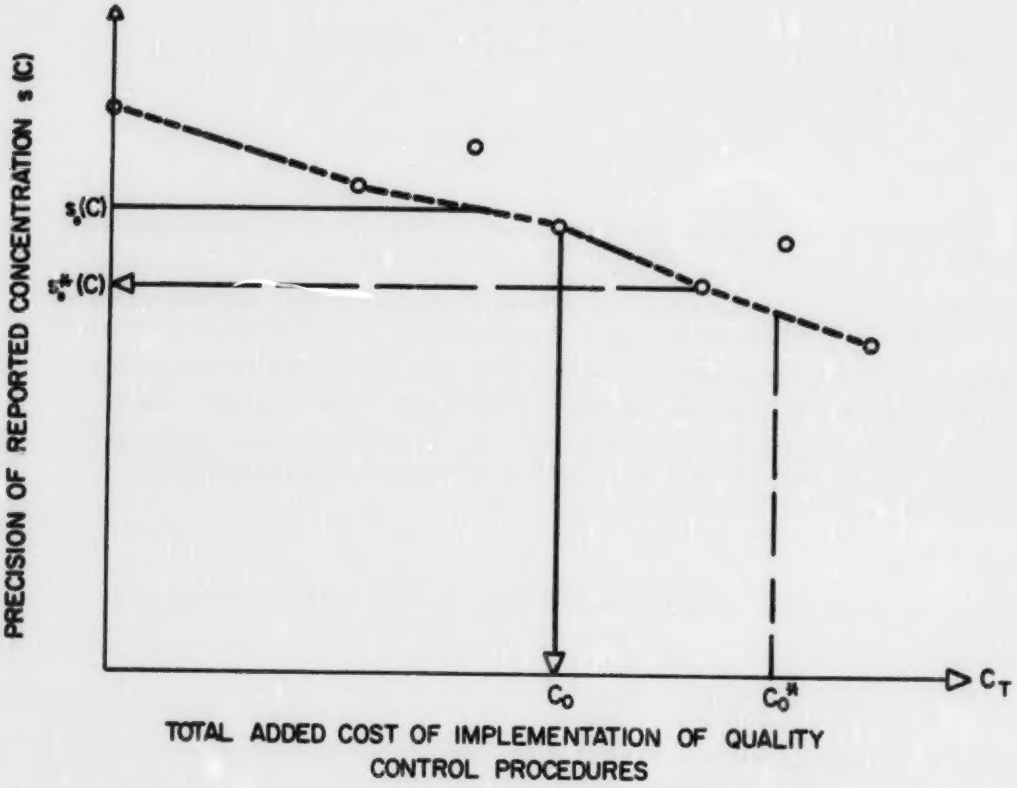


Figure 13: Precision of Reported C vs. Total Added Cost, C_T

DATA QUALITY ASSESSMENT

This section contains a discussion of the use of audit data to assess individual measurements and of techniques for the combination of these results to yield an overall assessment. The precision and biases of the individual measurements and operational procedures are first estimated. These results are then used to make the overall assessment as required in the second method or approach. The following section discusses a method of data presentation from a general point of view. This is followed by some specific statistical procedures to be employed in the assessment analysis.

Data Presentation

A reported value whose precision and accuracy (bias) are unknown is of little, if any, worth. The actual error of a reported value--that is, the magnitude and sign of its deviation from the true value--is usually unknown. Limits to this error, however, can usually be inferred, with some risk of being incorrect, from the precision of the measurement process by which the reported value was obtained and from reasonable limits to the possible bias of the measurement process. The bias, or systematic error, of a measurement process is the magnitude and direction of its tendency to measure something other than what was intended; its precision refers to the closeness or dispersion of successive independent measurements generated by repeated applications of the process under specified conditions, and its accuracy is determined by the closeness to the true value characteristic of such measurements.

Precision and accuracy are inherent characteristics of the measurement process employed and not of the particular end result obtained. From experience with a particular measurement process and knowledge of its sensitivity to uncontrolled factors, one can often place reasonable bounds on its likely systematic error (bias). This has been done in the model for the measured concentration as indicated in Table 37. It is also necessary to know how well the particular value in hand is likely to agree with other values that the same measurement process might have provided in this instance or might yield on measurements of the same magnitude on another occasion. Such information is provided by the estimated standard deviation of the reported value, which measures (or is an index of) the characteristic disagreement of repeated determinations of the same quantity by the same method and thus serves to indicate the precision (strictly, the imprecision) of the reported value.

A reported result should be qualified by a quasi-absolute type of statement that places bounds on its systematic error and a separate statement of its standard deviation, or of an upper bound thereto, whenever a reliable determination of such value is available. Otherwise, a computed value of the standard deviation should be given together with a statement of the number of degrees of freedom on which it is based.

As an example, consider strategy A0 in Table 37. Here, the assumed standard deviation and bias for a true SO_2 concentration of 380 $\mu g\ SO_2/m^3$ are $\sigma_T = 19\ \mu g\ SO_2/m^3$ (5.0% of 380 $\mu g/m^3$) and $\tau = -15\ \mu g\ SO_2/m^3$, respectively. The results would be reported as the measured concentration $(\mu g\ SO_2/m^3)_m$ minus the bias and with the following 2σ limits along with the audit level and lot size N; e.g.,

$$(\mu g\ SO_2/m^3)_m + 15 \pm 38,\ n = 7,\ N = 100.$$

For concentrations other than 380 $\mu g\ SO_2/m^3$, the overall standard deviation is obtained by

$$\sigma_T(\%) = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}\ , \text{ and}$$

$$\sigma_T(\mu g/m^3) = \sigma_T(\%) \times \mu g\ SO_2/m^3\ .$$

Assessment of Individual Measurements

The data collected during the audit program can also be used to assess the precision/accuracy of the individual measurements. For example, suppose that n = 7 independent audits are made of a particular measurement (e.g., the concentration of a calibration gas) and that it is desired to estimate the standard deviation of the reported data. In the first approach to auditing the data, the difference between the audited value and the original measured value by the operator or analyst was compared to the permitted deviation or suggested standard, and the measurement was classified as defective or non-defective on this basis. In this second approach the difference between the two corresponding measurements is treated quantitatively to obtain the standard deviation and bias.

Let d_j be the difference between the two measurements, say operator measurement less the audited measurement. Then

$$\bar{d} = \frac{\sum_{j=1}^{n} d_j}{n}$$

is the mean difference and a measure of bias in the particular measurement. The standard deviation of the differences is given by

$$s_d = \sqrt{\frac{\Sigma(d_j - \bar{d})^2}{n-1}}\ .$$

This is a measure of precision of the differences of two measurements. As described in Appendix 4, this can be converted to the standard deviation of a single measurement by dividing the result by $\sqrt{2}$ if each measurement can be assumed to have the same precision. If the audited measurement is made by a much more precise means than that of the routine data, it is not necessary to divide by $\sqrt{2}$ as the s_d will be almost equal to that of the original measurement.

The quantity s_d is also useful in checking for bias. The procedure is to compute

$$t = \frac{\overline{d}}{s_d}$$

and compare this value with the tabulated value of t for n-1 degrees of freedom (See Ref. 27), and if it is larger in absolute value than the t value for a preselected level of significance, then it is inferred that the mean difference is not due to chance but to some difference in the two instruments, operators, or conditions under which the measurements are taken, suggesting a bias in the resulting data.

The value of s_d or the corresponding standard deviation of an individual measurement $s_d/\sqrt{2}$ can be compared to the suggested standard as indicated in the subsection entitled Estimated Variance of Reported Data Using the Test Statistic $s^2/\sigma^2 = \chi^2/f$.

Overall Assessment of Data Quality

Suppose that the bias and standard deviation of each of m important variables in the measurement process can be obtained as suggested in the previous subsection. These results are then combined to obtain an overall estimate of the bias and standard deviation as follows. An estimate of the overall bias is given by

$$\hat{\tau} = \overline{d}_1 + \overline{d}_2 + \ldots + \overline{d}_m$$

if an additive relationship exists between the individual measurements and the desired measured concentration. In some cases a multiplicative relationship exists and percentage biases were added.

A similar approach is employed in combining the standard deviations of the individual measurements. The overall estimate

$$\hat{\sigma}_\tau = \sqrt{s_1^2 + s_2^2 + \ldots s_m^2}$$

where estimates of the standard deviations are obtained for m measurements. See Appendix 4 for the assumptions made in deriving the above results.

In modeling a complex relationship, computer programs exist for simulating the distribution of the measured concentration and for determining the sensitivity of the measurements to instrument errors, operator variation, and environmental effects. These are described in Appendices 1 and 2. With the use of these programs, errors/variations in the several process variables can be propagated through the use of the mathematical relationships to obtain the expected variation/bias in the measured concentration. This approach does not require making the assumptions concerning the additivity of the biases and variances as employed above.

In cases where biases are expected, it is usually desirable to combine the bias and the standard deviation to obtain a mean square error (MSE), actually the square root of the value, as follows:

$$\text{MSE} = \sqrt{(\text{bias})^2 + (\text{standard deviation})^2} \ ,$$

$$\text{MSE} = \sqrt{\hat{\tau}^2 + \sigma^2} \ .$$

This approach seems justified because the bias is expected to vary from one agency-laboratory to another resulting in a distribution of biases rather than all laboratories having the same bias. This laboratory bias is observed from examination of the collaborative test results, where very often a laboratory which obtains high test results on one day under given conditions continues to do so every day, for a laboratory obtaining low concentrations, all results tend to be low, even though both laboratories are measuring samples having the same concentration.

These measures of bias and precision are used in a cost versus precision analysis as described in the previous section under the heading "Cost Trade-Off Analyses." The methods of reporting or presenting the data are given in the subsection entitled Data Presentation.

Estimated Variance of Reported Data Using the Test Statistic: $s^2/\sigma^2 = \chi^2/f$

In estimating the variance of the measured concentration of a pollutant, the standard deviations of each of the variables are either hypothesized or estimated from available intra- and inter-laboratory tests. It is frequently desired to compare an estimated standard deviation based on field data with the assumed value based on limited test data for the purpose of determining if the true standard deviation is larger than the hypothesized value. In order to make this comparison, the ratio of the estimate s to the assumed value σ is obtained. If this ratio is larger than would be explained by chance, then corrective action should be taken to determine the assignable cause of the large value. It is also possible that the assumed value is too small because it was based on limited data or an engineering guess in some cases where no data are available.

Methods are given in each of the field documents for estimating the variance of a particular measurement, and the assumed variances (or standard deviations) are given. Thus, if one computes the ratio, s^2/σ^2, this value can be compared with a tabulated value of χ^2/f (read chi-square divided by degrees of freedom f = n-1, i.e., sample size less one) which is given in Ref. 27. If we are interested in detecting only unusually large values, then the critical value of χ^2/f may be taken as one exceeded only 5% or 10% of the time by chance. For example, if n=10, f=9, we read from the χ^2/f table that value exceeded 5% of the time (95 percent value) is 1.88. Hence, if s^2/σ^2 exceeds 1.88, one would infer that σ^2 is larger than the assumed value with 5% risk of being incorrect.

If on the other hand we are interested in detecting large deviations from the assumed value, i.e., unexpected large or small values, a two-sided test can be made. For example, if n=10, f=9, the two critical values are 0.30 and 2.114, outside of which 5% of the values fall. Hence, if the ratio s^2/σ^2 is less than 0.3 or larger than 2.114, one would infer that the variance is either smaller or larger than the assumed value with a 5% chance of being incorrect.

This same distribution can be employed to set up a control chart for the variance of a measurement of interest or in obtaining a confidence interval estimate of the true variance. In the latter case a 95% confidence interval estimate of σ^2 may be obtained by using the result that σ^2 falls within the interval,

$$\frac{s^2}{\left.\frac{\chi^2}{f}\right|_{.975}} \leq \sigma^2 \leq \frac{s^2}{\left.\frac{\chi^2}{f}\right|_{.025}}$$

$(\chi^2/f|_p$ denotes the p^{th} percentile of the distribution) 95% of the time. For example, if n=10, f=9, we have

$$\frac{s^2}{2.114} \leq \sigma^2 \leq \frac{s^2}{0.300}$$

with 95% confidence.

Critical values of χ^2/f vs. sample size are tabulated in Table 38 and plotted in Figure 14.

Table 38. CRITICAL VALUES OF s_i/σ_i

Level of Confidence	Statistic	Audit Level				
		n=5	n=10	n=15	n=20	n=25
90%	s_i/σ_i	1.40	1.28	1.23	1.20	1.18
95%	s_i/σ_i	1.54	1.37	1.30	1.26	1.23

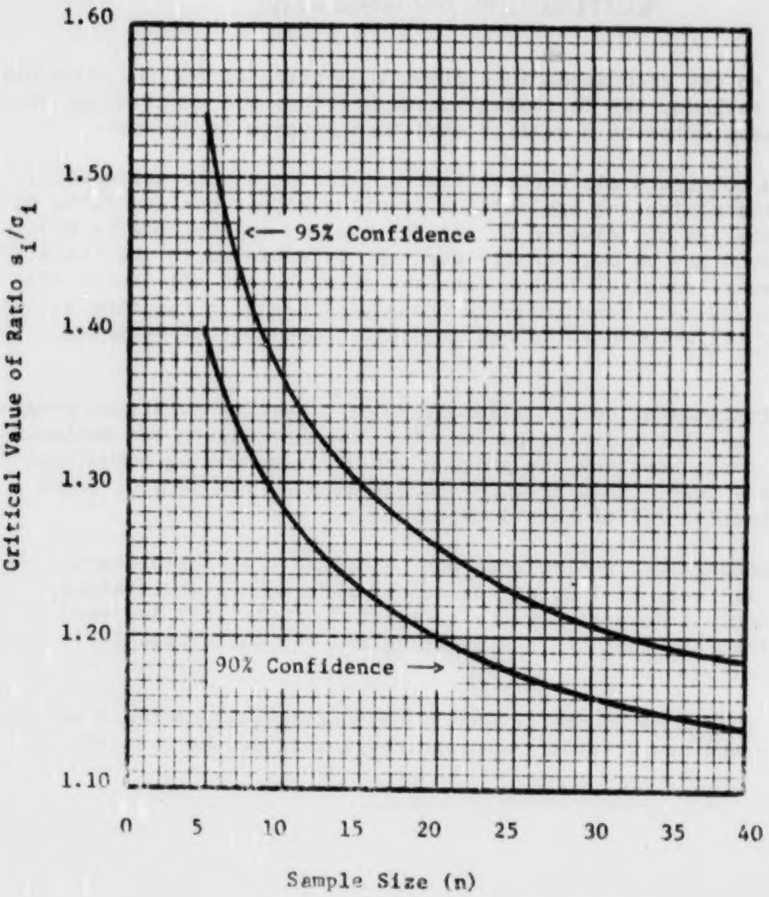


Figure 14: Critical Values of Ratio s_i/σ_i Vs. n

SECTION IV RECOMMENDATIONS AND CONCLUSIONS

In addition to the recommendations given in Section II for the individual measurement methods, the following recommendations and conclusions pertain to the implementation of a quality assurance program in general.

In the field documents all control limits were set at the estimated 3σ value. If the monitoring application requires higher quality data, the control limits can be adjusted to the 2σ level by multiplying by 2/3. Some limits as given may be more nearly what can be achieved in the field than others; however, once sufficient data are obtained and all limits are adjusted, the overall measurement process will begin to function in such a manner as to allow for a valid and accurate assessment of outgoing data quality.

Successful implementation and maintenance of a quality assurance program will require special training periods for the supervisors and managers responsible for conducting the program and decision making concerning data quality versus cost. This training should include the subject areas treated in Section III of this document.

Also, the managers and supervisors must converse with the operator. He should be: (1) well trained on the measurement method of interest, (2) continuously informed of what is expected of him, (3) informed of the quality of the data that he is generating and (4) encouraged to improve the quality of his work.

The present guideline documents should be continuously reviewed and updated as more data and new control techniques/equipment become available.

REFERENCES FINAL REPORT

1. Herbert C. McKee et al., "Collaborative Study of Reference Method for the Determination of Suspended Particulates in the Atmosphere (High Volume Method)," Southwest Research Institute, Contract CAP 70-40, SwRI Project 21-2811, San Antonio, Texas, June 1971.

2. Robert E. Lee, Jr. and Jack Wagman, "A Sampling Anomaly in the Determination of Atmospheric Sulfate Concentration," American Industrial Hygiene Association Journal 27, pp. 266-271, May-June 1966.

3. Robert M. Burton et al., "Field Evaluation of the High-Volume Particle Fractionating Cascade Impactor--A Technique for Respirable Sampling," presented at the 65th Annual Meeting of the Air Pollution Control Association, June 18-22, 1972.

4. Peter K. Mueller et al., "Selection of Filter Media: An Annotated Outline," presented at the 13th Conference on Methods in Air Pollution and Industrial Hygiene Studies, University of California, Berkeley, California, October 30-31, 1972.

5. G. P. Tierney and W. D. Conner, "Hygroscopic Effects on Weight Determinations of Particulates Collected on Glass-Fiber Filters," American Industrial Hygiene Association Journal 28, pp. 363-365, July-August, 1967.

6. John F. Kowalczyk, "The Effects of Various Pre-Weighing Procedures on the Reported Weights of Air Pollutants Collected by Filteration," presented at the 60th Annual Meeting of the Air Pollution Control Association, Cleveland, Ohio, June 11-16, 1967.

7. C. D. Robson and K. E. Foster, "Evaluation of Air Particulate Sampling Equipment," American Industrial Hygiene Association Journal 23, pp. 404-410, 1962.

8. John S. Henderson, "A Continuous-Flow Recorder for the High-Volume Air Sampler," presented at the 8th Conference on Methods in Air Pollution and Industrial Hygiene Studies, Oakland, California, February 6-8, 1967.

9. Walter K. Harrison et al., "Constant Flow Regulators for the High-Volume Air Sampler," American Industrial Hygiene Association Journal 21, pp. 115-120, 1960.

10. Franklin Smith and A. Carl Nelson, "Guidelines for Development of a Quality Assurance Program for Reference Method For the Determination of Suspended Particulates In The Atmosphere (High Volume Method)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, April 1973.

11. H. C. McKee et al., "Collaborative Study of Reference Method for Determination of Sulfur Dioxide in the Atmosphere (Pararosaniline Method)," Southwest Research Institute, Contract CPA-70-40, SwRI Project 21-2811, San Antonio, Texas, September 1971.

12. J. P. Lodge, Jr., et al., "The Use of Hypodermic Needles as Critical Orifices in Air Sampling," Journal of the Air Pollution Control Association 16 (4), April 1966, pp. 197-200.

13. F. P. Scaringelli, B. E. Saltzman, and S. A. Frey, "Spectrophotometric Determination of Atmospheric Sulfur Dioxide," Analytical Chemistry 39, page 1709, December 1967.

14. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program for Reference Method For Determination of Sulfur Dioxide In the Atmosphere (Pararosaniline Method)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, June 1973.

15. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program For Reference Method for the Measurement of Photochemical Oxidants Corrected for Interferences Due to Nitrogen Oxides and Sulfur Dioxide," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, April 1973.

16. Herbert C. McKee, and Ralph S. Childers, Provisional Report of Collaborative Study of Reference Method for Measurement of Photochemical Oxidants Corrected for Interference Due to Nitrogen Oxides and Sulfur Dioxide, Contract CPA 70-40.

17. J. A. Hodgeson, R. K. Stevens, and B. E. Martin, "A Stable Ozone Source Applicable as a Secondary Standard for Calibration of Atmospheric Monitors," Air Quality Instruments Vol. 1, Instrument Society of America, Pittsburg, 1972, pp. 149-158.

18. "Tentative Method For the Continuous Measurement of Nitrogen Dioxide (Chemiluminescent)," Federal Register Vol. 38, No. 110, Part II, Friday, June 8, 1973, pages 15177-15180.

19. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program for the Continuous Measurement of Nitrogen Dioxide in the Ambient Air (Chemiluminescent)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, June, 1973.

20. Franklin Smith and A. Carl Nelson, Jr., "Guidelines for Development of a Quality Assurance Program for Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, March 1973.

11. H. C. McKee et al., "Collaborative Study of Reference Method for Determination of Sulfur Dioxide in the Atmosphere (Pararosaniline Method)," Southwest Research Institute, Contract CPA-70-40, SwRI Project 21-2811, San Antonio, Texas, September 1971.

12. J. P. Lodge, Jr., et al., "The Use of Hypodermic Needles as Critical Orifices in Air Sampling," Journal of the Air Pollution Control Association 16 (4), April 1966, pp. 197-200.

13. F. P. Scaringelli, B. E. Saltzman, and S. A. Frey, "Spectrophotometric Determination of Atmospheric Sulfur Dioxide," Analytical Chemistry 39, page 1709, December 1967.

14. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program for Reference Method For Determination of Sulfur Dioxide In the Atmosphere (Pararosaniline Method)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, June 1973.

15. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program For Reference Method for the Measurement of Photochemical Oxidants Corrected for Interferences Due to Nitrogen Oxides and Sulfur Dioxide," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, April 1973.

16. Herbert C. McKee, and Ralph S. Childers, Provisional Report of Collaborative Study of Reference Method for Measurement of Photochemical Oxidants Corrected for Interference Due to Nitrogen Oxides and Sulfur Dioxide, Contract CPA 70-40.

17. J. A. Hodgeson, R. K. Stevens, and B. E. Martin, "A Stable Ozone Source Applicable as a Secondary Standard for Calibration of Atmospheric Monitors," Air Quality Instruments Vol. 1, Instrument Society of America, Pittsburg, 1972, pp. 149-158.

18. "Tentative Method For the Continuous Measurement of Nitrogen Dioxide (Chemiluminescent)," Federal Register Vol. 38, No. 110, Part II, Friday, June 8, 1973, pages 15177-15180.

19. Franklin Smith and A. Carl Nelson, Jr., "Guidelines For Development of a Quality Assurance Program for the Continuous Measurement of Nitrogen Dioxide in the Ambient Air (Chemiluminescent)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, June, 1973.

20. Franklin Smith and A. Carl Nelson, Jr., "Guidelines for Development of a Quality Assurance Program for Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry)," Research Triangle Institute, Contract EPA-Durham 68-02-0598, RTI Project 43U-763, Research Triangle Park, North Carolina, March 1973.

21. Herbert C. McKee et al., "Collaborative Study of Reference Method for the Continuous Measurement of Carbon Monoxide in the Atmosphere (Non-Dispersive Infrared Spectrometry)," Southwest Research Institute, Contract CPA 70-40, SwRI Project 01-2811, San Antonio, Texas, May 1972.

22. John Mandel, The Statistical Analysis of Experimental Data, Interscience Publishers, Division of John Wiley & Sons, New York, N. Y., 1964.

23. Anderson, R. L. and Bancroft, T. A., Statistical Theory in Research McGraw-Hill Book Company, Inc., New York, 1952.

24. C. A. Bennett and N. L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, John Wiley & Sons, Inc., New York, 1954.

25. Bartholomew P. Hsi, "Optimization of Quality Control in the Chemical Laboratory," Technometrics Vol. 8, No. 3, August 1966, pp. 519-534.

26. A. Hald, Statistical Theory With Engineering Applications, John Wiley & Sons, Inc., New York, 1952.

27. A. Hald, Statistical Tables and Formulas, John Wiley and Sons, Inc., New York, 1952.

28. W. A. Wallis, "Tolerance Intervals for Linear Regression," Second Berkeley Symposium.

29. W. A. Shewhart, Economic Control of Quality of Manufactured Product, D. Van Nostrand Company, Inc., 1931.

30. S. W. Roberts, "A Comparison of Some Control Chart Procedures," Technometrics, Vol. 8, No. 3, August 1966, pp. 411-430.

APPENDIX I COMPUTER SIMULATION

GENERAL

Performance models were developed and utilized in writing guideline documents for the reference methods for measuring suspended particulates and sulfur dioxide. Simulation analyses were performed for both methods using the performance models. The simulation process is described using the high volume method as an example with sample computer printouts of the results. Only sample computer printouts are presented for the pararosaniline method. For details on the performance models or individual variable models, refer to the appropriate subsection of **Section II of this document.**

SIMULATION OF THE HIGH VOLUME METHOD OF MEASURING S.P.

In order to estimate how the concentration of particulate matter in air varies as a function of the errors and/or variations in weighing, calibration, reading flow rates, "pH of filter," relative humidity, etc., two types of analyses are performed. They are a simulation analysis as described in this Appendix and a sensitivity analysis as described in Appendix 2. In the simulation analysis the particulate concentration is denoted by S.P. and expressed as a function of thirteen pertinent variables denoted by X(i), i = 1, ..., 13. These variables and the estimates of their mean values and variations are required to perform the simulation. For example, for a normally distributed variable its nominal or mean value is **given along with its standard deviation (see Appendix 3 for a discussion of the normal distribution).**

Some of the variables of the model for particulate concentration were assumed to be uniformly distributed, i.e., all deviations from their nominal values were considered to be equally likely over the range of variation **indicated (see Appendix 3 for a discussion of the uniform distribution.) All other variables were assumed to be normally distributed. This input information is given in Table 1-1.**

The next step in the process is to generate the random numbers with the specified distributions and insert them into the model to obtain simulated values of S.P. A sample of the first few values of the 100 values generated is given in Table 1-2. In Table 1-3 is a printout of the computed means and standard deviations of each of the variables in order to provide a gross check on the simulated values. In Table 1-4 is given a ranking of the 100 simulated S.P. values in ascending order.

In summary, there are certain significant results to be derived from the tabulated results. One is that the values of the S.P.'s shown in ascending order can be plotted on normal probability paper (see Figure 1-1) to indicate the expected variation of the concentration as a function of the accumulation of errors of analytical method. A normal distribution with the mean and standard deviation computed from the simulated values is also shown on Figure 1-1 for comparison with the simulated distribution. The moments as computed and given in Table 1-5 can be used to test the deviation from normality if it is desired. The standard deviation, 7.1 $\mu g/m^3$, is 7.1% of the mean value.

The simulation analysis is clearly no better than the model and the estimated values of the input data. These inputs are based on some actual inter-laboratory tests and intra-laboratory tests; and when data are not available, they are based on engineering and statistical judgments.

TABLE 1-1: Input Data

MODEL 1. SE01	VAR. NAMES	NOMINAL VALUE	DEVIATION	DISTRIBUTION
1	X1	.65000E 1	.60000E 1	UNIFORM*
2	X2	.55000E 2	.50000E 1	NORMAL**
3	X3	.00000E 0	.10000E -2	NORMAL
4	X4	-.50000E -1	.10000E 0	UNIFORM
5	X5	-.20000E -1	.12000E 0	UNIFORM
6	X6	.00000E 0	.50000E -1	NORMAL
7	X7	.00000E 0	.30000E -1	NORMAL
8	X8	.00000E 0	.20000E -1	NORMAL
9	X9	-.70000E -1	.14000E 0	UNIFORM
10	X10	.10000E 1	.20000E 2	UNIFORM
11	X11	.20000E 2	.50000E 1	NORMAL
12	X12	.00000E 0	.50000E 1	NORMAL
13	X13	.80000E 0	.84000E 0	UNIFORM

* For all uniformly distributed variables the lowest value (6.5 in this case) and the range of the distribution (6 in this case) are given; thus X(1) ranges from 6.5 to 12.5, and all deviations in this interval are equally likely.

** For all normally distributed variables the mean (nominal in table) and standard deviation are given. Thus X(2) has mean 55 and a standard deviation of 5.

TABLE 1-2: Portion of Simulated Data

X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	X(9)	X(10)	X(11)	X(12)	X(13)	S.P.
.7830E 1	.4782E 2	.1218E -2	.3853E -1	.1152E -1	.3651E -1	.2990E -3	-.1627E -	.5093E -1	.1207E 2	.1629E 2	-.6551E 1	.8775E 0	.9937E 2
.7068E 1	.5970E 2	-.1612E -2	.7080E -2	.8629E -1	.2536E -6	.4795E -1	.2783E -1	.1853E -1	.1932E 2	.1655E 2	.1609E 1	.9981E 0	.8876E 2
.1149E 2	.6284E 2	.6437E -4	-.4106E -1	.9977E -1	-.5824E -1	-.4985E -2	-.1075E -2	-.4300E -1	.9281E 1	.1654E 2	-.2220E 1	.1259E 1	.1101E 3
.6980E 1	.5816E 2	-.5803E -3	-.3501E -1	.9953E -1	-.1802E -1	.5694E -2	.3536E -1	-.1415E -1	.5277E 1	.3435E 2	-.9324E 1	.9070E 0	.9421E 2
.8029E 1	.5002E 2	.4584E -4	-.3940E -1	.4551E -1	.6223E -1	.3764E -1	.3962E -1	-.2317E -1	.2738E 1	.1904E 2	.1842E 1	.1483E 1	.9083E 2
.1177E 2	.5432E 2	-.2776E -2	.4150E -2	.9027E -1	-.5874E -1	.2293E -1	.5693E -2	.6692E -1	.7484E 1	.2554E 2	.2328E -1	.1498E 1	.1040E 3
.9441E 1	.5119E 2	-.1662E -2	.4351E -1	.8875E -1	-.2245E -1	-.7843E -2	-.3004E -2	-.2960E -1	.1245E 2	.2130E 2	-.1314E 1	.1549E 1	.9774E 2
.9430E 1	.4680E 2	.9699E -3	.2456E -1	.5852E -1	.1285E -1	.2722E -1	.1034E -1	-.1825E -1	.5941E 1	.2122E 2	-.7667E 0	.1092E 1	.9099E 2
.9729E 1	.4610E 2	-.3780E -3	.7666E -2	-.1051E -1	-.2684E -1	.1140E -1	-.1659E -1	-.6228E -1	.1840E 2	.2036E 2	.3658E 1	.1458E 1	.1011E 3
.1197E 2	.6340E 2	-.1214E -2	.2622E -1	.4258E -2	.1642E -1	-.2222E -1	-.1173E -3	.1032E -1	.2065E 2	.1932E 2	-.6009E 1	.1369E 1	.1019E 3
.8891E 1	.5369E 2	-.1603E -2	-.4692E -1	-.1227E -1	-.4097E -1	.3728E -1	-.2776E -1	.1880E -1	.2061E 2	.1439E 2	.3702E 1	.1210E 1	.1102E 3
.7379E 1	.5580E 2	.4721E -3	.9131E -2	.4750E -1	.3343E -2	.2967E -1	.9069E -3	.1265E -1	.1161E 2	.2167E 2	.3372E 1	.1488E 1	.9590E 2
.6928E 1	.5932E 2	-.5014E -3	-.2026E -1	.8348E -1	.6269E -1	-.5143E -2	-.2311E -1	-.6638E -1	.1907E 2	.1785E 2	.1723E 1	.8037E 0	.8616E 2
.7666E 1	.5277E 2	.1159E -2	-.2671E -1	.6824E -1	-.7494E -1	.7235E -2	-.1628E -1	-.1128E -1	.1881E 2	.1932E 2	-.5842E 1	.1449E 1	.1109E 3
.9840E 1	.5237E 2	.8311E -3	-.1235E -1	.3672E -1	-.5127E -1	-.2259E -1	.1103E -1	-.3780E -1	.1377E 2	.2372E 2	.9482E 1	.1080E 1	.1044E 3

TABLE 1-3 Check of Simulated Values

```
INPUT CHECK
                                           Mean            Standard
MODEL  1, SEQ1         VAR, NAMES    NOMINAL VALUE        DEVIATION     DISTRIBUTION
                  1        X1          .94695E  1        .17188E  1     UNIFORM*
                  2        X2          .54879E  2        .50161E  1     NORMAL
                  3        X3         -.93897E -4        .10161E -2     NORMAL
                  4        X4          .46280E -3        .28897E -1     UNIFORM
                  5        X5          .40245E -1        .35160E -1     UNIFORM
                  6        X6          .15023E -2        .48554E -1     NORMAL
                  7        X7          .14006E -2        .30998E -1     NORMAL
                  8        X8          .17010E -2        .19365E -1     NORMAL
                  9        X9         -.20617E -2        .40943E -1     UNIFORM
                 10        X10         .11067E  2        .57198E  1     UNIFORM
                 11        X11         .20141E  2        .45560E  1     NORMAL
                 12        X12        -.23813E  0        .55395E  1     NORMAL
                 13        X13         .12168E  1        .23924E  0     UNIFORM
```

---

*In the check the means and standard deviations are given for each variable. For X(1) the mean is 9.47 and the standard deviation is 1.72; these values can be readily compared to the expected mean and standard deviation of a uniform variable on the interval 6.5 to 12.5, which would have a mean of 9.5 and a standard deviation $\sqrt{3} = 1.73$. This is unusually good agreement.

TABLE 1-4: Listing of Simulated Values in Ascending Order

I	I/N	S.P.
1	.010	.8416E 2
2	.020	.8616E 2
3	.030	.8731E 2
4	.040	.8876E 2
5	.050	.8892E 2
6	.060	.8923E 2
7	.070	.9034E 2
8	.080	.9049E 2
9	.090	.9069E 2
10	.100	.9083E 2
11	.110	.9099E 2
12	.120	.9116E 2
13	.130	.9146E 2
14	.140	.9226E 2
15	.150	.9240E 2
16	.160	.9317E 2
17	.170	.9339E 2
18	.180	.9386E 2
19	.190	.9409E 2
20	.200	.9417E 2
21	.210	.9421E 2
22	.220	.9426E 2
23	.230	.9456E 2
24	.240	.9463E 2
25	.250	.9493E 2
26	.260	.9522E 2
27	.270	.9554E 2
28	.280	.9557E 2
29	.290	.9558E 2
30	.300	.9590E 2
31	.310	.9626E 2
32	.320	.9660E 2
33	.330	.9665E 2
34	.340	.9686E 2
35	.350	.9719E 2
36	.360	.9729E 2
37	.370	.9734E 2
38	.380	.9736E 2
39	.390	.9752E 2
40	.400	.9774E 2
41	.410	.9784E 2
42	.420	.9786E 2
43	.430	.9788E 2
44	.440	.9795E 2
45	.450	.9797E 2
46	.460	.9834E 2
47	.470	.9853E 2
48	.480	.9854E 2
49	.490	.9860E 2
50	.500	.9866E 2
51	.510	.9885E 2
52	.520	.9936E 2
53	.530	.9937E 2
54	.540	.9937E 2
55	.550	.9954E 2
56	.560	.9975E 2
57	.570	.1002E 3
58	.580	.1005E 3
59	.590	.1006E 3
60	.600	.1008E 3
61	.610	.1009E 3
62	.620	.1009E 3
63	.630	.1011E 3
64	.640	.1017E 3
65	.650	.1019E 3
66	.660	.1021E 3
67	.670	.1026E 3
68	.680	.1032E 3
69	.690	.1033E 3
70	.700	.1038E 3
71	.710	.1040E 3
72	.720	.1040E 3
73	.730	.1042E 3
74	.740	.1044E 3
75	.750	.1044E 3
76	.760	.1047E 3
77	.770	.1051E 3
78	.780	.1053E 3
79	.790	.1063E 3
80	.800	.1065E 3
81	.810	.1072E 3
82	.820	.1074E 3
83	.830	.1074E 3
84	.840	.1076E 3
85	.850	.1078E 3
86	.860	.1079E 3
87	.870	.1083E 3
88	.880	.1086E 3
89	.890	.1086E 3
90	.900	.1090E 3
91	.910	.1096E 3
92	.920	.1100E 3
93	.930	.1101E 3
94	.940	.1102E 3
95	.950	.1109E 3
96	.960	.1136E 3
97	.970	.1143E 3
98	.980	.1146E 3
99	.990	.1179E 3
100	1.000	.1196E 3

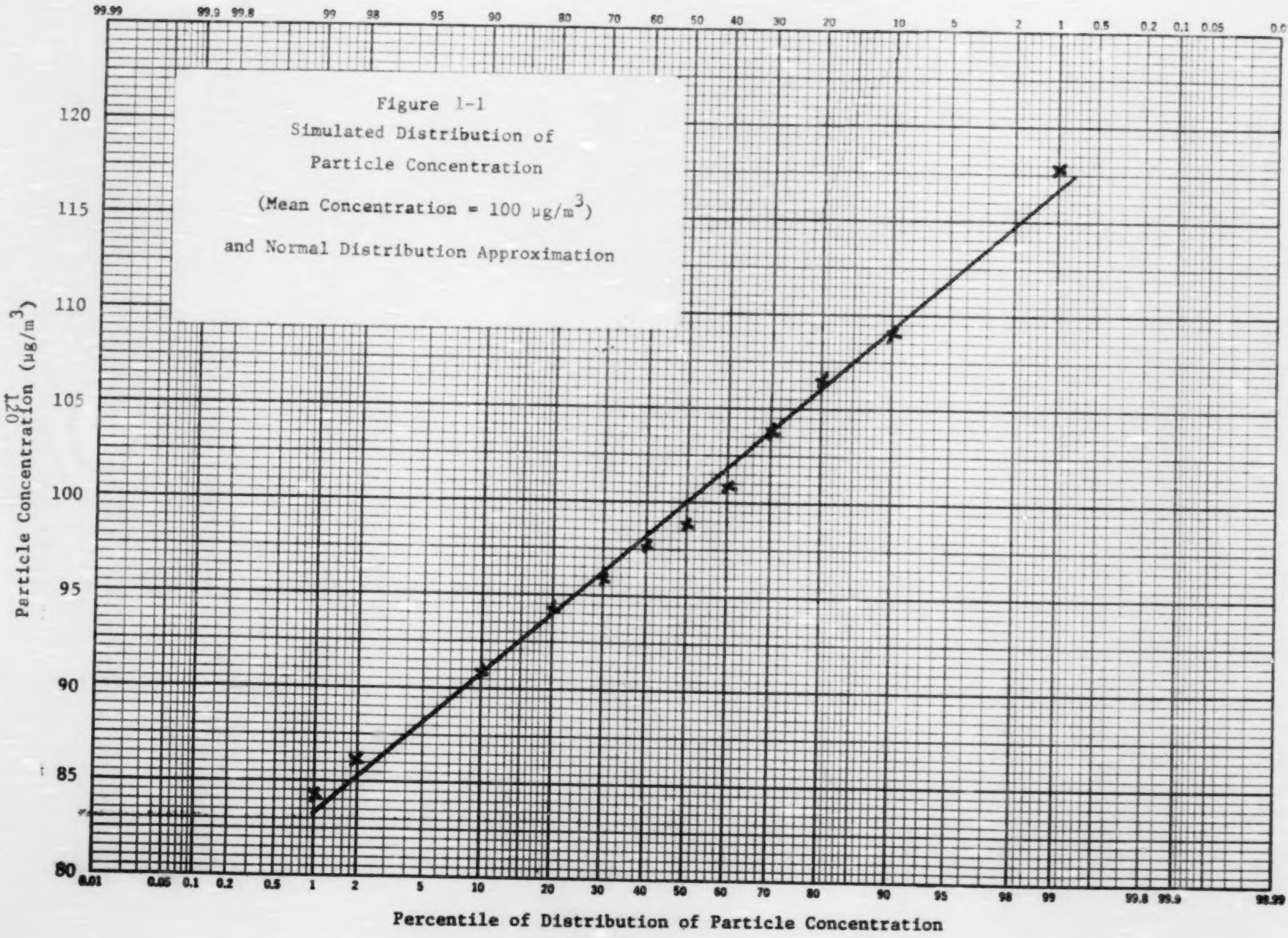


Figure 1-1
Simulated Distribution of
Particle Concentration
(Mean Concentration = 100 μg/m^3)
and Normal Distribution Approximation

TABLE 1-5: **Moments of Simulated Distribution**

```
MOMENTS          S.P.

      FIRST     999661E  2

     SECOND    .502728E  4

      THIRD    .127074E  5

     FOURTH    .723892E  6

  STD. DEV.    .712604E  1

   SKEWNESS    .356498E  0

   KURTOSIS    .286422E -1

VARIANCE - COVARIANCE MATRIX, ORDER 3

VARIANCE       .515757E  2
```

SIMULATION OF THE PARAROSANILINE METHOD OF MEASURING SO_2

Simulation analysis of the pararosaniline method of measuring SO_2 was performed in the same manner as described in the preceeding section of this appendix for the High Volume Method. The SO_2 concentration is denoted by Y and expressed as a function of fifteen variables denoted by X(i), i = 1, 2, ..., 15. These variables and the estimates of their mean values are required to perform the simulation. For example, for a normally distributed variable its nominal or mean value is given along with its standard deviation (see Appendix 3 for a discussion of the normal distribution).

Some of the variables of the model for SO_2 concentration in $\mu g SO_2/m^3$ were assumed to be uniformly distributed; i.e., all deviations from their nominal values were considered to be equally likely over the range of variation indicated (see Appendix 3 for a discussion of the uniform distribution). All other variables were assumed to be normally distributed. This input information is given in Table 1-6.

The next step in the process is to generate the random numbers with the specified distributions and insert them into the model to obtain simulated values of SO_2. A sample of the first few values of the 100 values generated is given in Table 1-7. In Table 1-8 is a printout of the computed means and standard deviations of each of the variables in order to provide a gross check on the simulated values. In Table 1-9 is given a ranking of the 100 simulated SO_2 values in ascending order.

In summary, there are certain significant results to be derived from the tabulated results. One is that the concentration values of SO_2 shown in ascending order can be plotted on normal probability paper (see Figure 1-1 as an example) to indicate the expected variation of the concentration as a function of the accumulation of errors of analytical method. A normal distribution with the mean and standard deviation computed from the simulated values was also shown on Figure 1-1 for comparison with the simulated distribution. Also, the moments of the distribution as computed and given in Table 1-10 can be used to test the deviation from normality if it is desired. The standard deviation, 23.8 $\mu g/m^3$, is 26% of the mean value. Results from a collaborative test of the method (Ref. 11) showed a standard deviation of 17 and 33 percent of the mean for repeatibility and reproducibility, respectively, at a true concentration of 100 $\mu g/m^3$.

The simulation analysis is clearly no better than the model and the estimated values of the input data. These inputs are based on some actual inter-laboratory tests and intra-laboratory tests; and when data are not available, they are based on engineering and statistical judgments.

TABLE 1-6: Input Data (SO_2)

MODEL 1, SE01	VAR. NAMES	NOMINAL VALUE	DEVIATION	DISTRIBUTION
1	X1	.30000E -1	.82000E -3	NORMAL[*]
2	X2	.16300E 0	.12000E -1	NORMAL
3	X3	.00000E 0	.12000E -1	NORMAL
4	X4	.00000E 0	.15000E -1	NORMAL
5	X5	.10000E 2	.28000E 0	NORMAL
6	X6	.00000E 0	.20000E -1	NORMAL
7	X7	-.40000E -1	.40000E -1	UNIFORM[**]
8	X8	-.40000E -1	.40000E -1	UNIFORM
9	X9	.00000E 0	.11200E -1	NORMAL
10	X10	.00000E 0	.57000E -2	NORMAL
11	X11	.00000E 0	.70000E 1	NORMAL
12	X12	-.50000E -1	.50000E -1	UNIFORM
13	X13	.15000E 2	.50000E 1	NORMAL
14	X14	.00000E 0	.10000E 2	UNIFORM
15	X15	-.50000E -1	.50000E -1	UNIFORM

[*] For all normally distributed variables the mean (nominal in table) and standard deviation are given. Thus X(1) has a mean of 0.03 and a standard deviation of 0.00082.

[**] For all uniformally distributed variables the lower limit and the range of the distribution are given. Variable X(7) has a lower limit of -0.04 and a range of 0.04.

TABLE 1-7: Portion of Simulated Data (SO_2)

```
         X1          X2          X3          X4          X5          X6          X7          X8
     X9         X10         X11         X12         X13         X14         X15        SEQ1

       .2870E -1   .1696E  0  -.1053E -1   .4841E -3   .1024E  2  -.2127E -1  -.2365E -1  -.1855E -1
  -.5548E -2  -.7000E -3   .4493E  1  -.3274E -1   .8906E  1   .4414E  1  -.1921E -1   .8279E  2

       .3037E -1   .1791E  0   .9943E -2   .8359E -2   .9827E  1   .3630E -1  -.3906E -1  -.7090E -2
  -.1121E -1  -.8403E -2   .1911E  1  -.2544E -1   .1091E  2   .2593E  1  -.1973E -1   .9844E  2

       .3213E -1   .1694E  0   .3374E -2  -.3105E -1   .1014E  2  -.2085E -1  -.3135E -1  -.3250E -1
   .1250E -1  -.7287E -2   .4495E  1  -.3923E -1   .1078E  2   .1553E  1  -.4133E -1   .1353E  3

       .2944E -1   .1660E  0   .1007E -1   .2143E -2   .1009E  2  -.3422E -2  -.2051E -1  -.3479E -1
  -.9447E -2  -.6694E -2  -.7867E  1  -.4912E -1   .1557E  2   .6294E  1  -.9033E -2   .9845E  2

       .3165E -1   .1523E  0  -.4246E -2   .6387E -2   .1008E  2   .4089E -2  -.2654E -1  -.3395E -1
  -.2781E -1  -.1394E -2   .5798E  1  -.3010E -1   .2593E  2   .1016E  1  -.4783E -1   .7403E  2

       .3099E -1   .1631E  0  -.3562E -2   .3205E -1   .1047E  2  -.4307E -1  -.2945E -1  -.9980E -2
   .6913E -2  -.9389E -3  -.6098E  1  -.7178E -2   .1558E  2   .3120E  1  -.3271E -1   .5414E  2

       .3062E -1   .1750E  0  -.3708E -3   .6285E -2   .9908E  1  -.1608E -1  -.9238E -2  -.2289E -1
  -.2110E -1   .1302E -2   .2831E  1  -.5347E -2   .1283E  2   .5285F  1  -.3870E -1   .8885E  2

       .3066E -1   .1529E  0   .3719E -2  -.8507E -2   .1030E  2   .1537E -1  -.2590E -1  -.1268E -1
  -.1010E -1  -.2768E -2   .3145E  1  -.4961E -1   .5408E  1   .3071E  1  -.4077E -1   .1167E  3

       .2925E -1   .1419E  0  -.1877E -1   .3629E -2   .9818E  1  -.8235E -2  -.1943E -1  -.3934E -1
   .4185E -3  -.5784E -2   .3006E  0  -.1497E -1   .1443E  2   .1719E  1  -.1394E -1   .6289E  2

       .3044E -1   .1653E  0  -.2218E -2  -.8488E -2   .9696E  1  -.3320E -1  -.9844E -2  -.2871E -2
  -.9800E -2  -.3594E -3  -.2383E  1  -.2642E -1   .1613E  2   .6147E  1  -.3320E -1   .9678E  2
```

TABLE 1-8: Check of Simulated Values (SO_2)

```
INPUT CHECK

MODEL  1. SEQ1
                        VAR. NAMES     NOMINAL VALUE       DEVIATION      DISTRIBUTION
                 1         X1           .30103E -1        .90222E -3      NORMAL*
                 2         X2           .16306E  0        .10523E -1      NORMAL
                 3         X3          -.19097E -2        .10127E -1      NORMAL
                 4         X4           .96563E -3        .15640E -1      NORMAL
                 5         X5           .99832E  1        .29516E  0      NORMAL
                 6         X6          -.19044E -2        .20955E -1      NORMAL
                 7         X7          -.20329E -1        .10525E -1      UNIFORM
                 8         X8          -.19719E -1        .12615E -1      UNIFORM
                 9         X9          -.12091E -2        .10373E -1      NORMAL
                10         X10          .40084E -3        .65102E -2      NORMAL
                11         X11         -.63892E  0        .72386E  1      NORMAL
                12         X12         -.26241E -1        .13890E -1      UNIFORM
                13         X13          .14678E  2        .54267E  1      NORMAL
                14         X14          .46915E  1        .30703E  1      UNIFORM
                15         X15         -.23934E -1        .14835E -1      UNIFORM
```

---

*In the check the means and standard deviations are given for each variable. For X(1) the mean or nominal value is 0.03 and the standard deviation is 0.0009. This is checked against the input data of Table 1-6 which shows a mean of 0.03 and a standard deviation of 0.00082 for X(1). The agreement is sufficient to accept the simulation data for X(1) as valid.

TABLE 1-9: Listing of Simulated Values in Ascending Order (SO_2)

```
DEPENDENT DATA LISTED IN ASCENDING ORDER

  I     I/N       SE01
  1     .010     .2980E  2       45     .450     .8828E  2        90     .900     .1229E  3
  2     .020     .4149E  2       46     .460     .8855E  2        91     .910     .1235E  3
  3     .030     .4645E  2       47     .470     .8885E  2        92     .920     .1237E  3
  4     .040     .4843E  2       48     .480     .8987E  2        93     .930     .1260E  3
  5     .050     .4994E  2       49     .490     .9115E  2        94     .940     .1264E  3
  6     .060     .5003E  2       50     .500     .9125E  2        95     .950     .1286E  3
  7     .070     .5326E  2       51     .510     .9129E  2        96     .960     .1304E  3
  8     .080     .5414E  2       52     .520     .9257E  2        97     .970     .1324E  3
  9     .090     .5880E  2       53     .530     .9289E  2        98     .980     .1353E  3
 10     .100     .6209E  2       54     .540     .9292E  2        99     .990     .1392E  3
 11     .110     .6347E  2       55     .550     .9494E  2       100    1.000     .1476E  3
 12     .120     .6397E  2       56     .560     .9495E  2
 13     .130     .6403E  2       57     .570     .9622E  2
 14     .140     .6404E  2       58     .580     .9678E  2
 15     .150     .6575E  2       59     .590     .9767E  2
 16     .160     .6583E  2       60     .600     .9844E  2
 17     .170     .6626E  2       61     .610     .9845E  2
 18     .180     .6893E  2       62     .620     .9889E  2
 19     .190     .6979E  2       63     .630     .9927E  2
 20     .200     .6998E  2       64     .640     .9999E  2
 21     .210     .7213E  2       65     .650     .1001E  3
 22     .220     .7307E  2       66     .660     .1003E  3
 23     .230     .7339E  2       67     .670     .1009E  3
 24     .240     .7399E  2       68     .680     .1027E  3
 25     .250     .7403E  2       69     .690     .1031E  3
 26     .260     .7645E  2       70     .700     .1037E  3
 27     .270     .7739E  2       71     .710     .1037E  3
 28     .280     .7824E  2       72     .720     .1054E  3
 29     .290     .7932E  2       73     .730     .1067E  3
 30     .300     .8097E  2       74     .740     .1068E  3
 31     .310     .8138E  2       75     .750     .1070E  3
 32     .320     .8158E  2       76     .760     .1070E  3
 33     .330     .8270E  2       77     .770     .1091E  3
 34     .340     .8279E  2       78     .780     .1107E  3
 35     .350     .8313E  2       79     .790     .1119E  3
 36     .360     .8380E  2       80     .800     .1133E  3
 37     .370     .8451E  2       81     .810     .1152E  3
 38     .380     .8535E  2       82     .820     .1157E  3
 39     .390     .8605E  2       83     .830     .1167E  3
 40     .400     .8622E  2       84     .840     .1188E  3
 41     .410     .8666E  2       85     .850     .1197E  3
 42     .420     .8689E  2       86     .860     .1201E  3
 43     .430     .8773E  2       87     .870     .1203E  3
 44     .440     .8797E  2       88     .880     .1214E  3
                                 89     .890     .1220E  3
```

TABLE 1-10: Moments of Simulated Distribution (SO_2)

```
MOMENTS

       FIRST    .918318F  2

      SECOND    .561767F  5

       THIRD   -.135215F  6

      FOURTH    .834260F  8

   STD. DEV.    .238210F  2

    SKEWNESS   -.101552F  0

    KURTOSIS    .264356E -1

   VARIANCE - COVARIANCF MATRIX, ORDER 1

   VARIANCE     .568121F  3
```

APPENDIX 2 SENSITIVITY ANALYSIS

GENERAL

Performance models were developed and utilized in writing the guideline documents for the reference methods for measuring suspended particulates and sulfur dioxide. Sensitivity analyses were performed for both methods using the performance models. The sensitivity analysis process is described using the High Volume Method as an example with sample computer printouts of the results. Computer printout and a short discussion of the results are presented for the pararosaniline method. For details on the performance models and individual variables, refer to the appropriate subsection of Section II of this document.

SENSITIVITY ANALYSIS OF THE HIGH VOLUME METHOD OF MEASURING SUSPENDED PARTICULATES

A sensitivity analysis is performed on the model using as input the mean (X) and deviation (DX) of each of the variables as given in the last two columns of the upper part of Table 2-1. Note that these inputs are identical to those used in the simulation analysis for the normal variables; however, in the case of the uniform variables the mean and 0.25 × range are used as inputs. The computer program then calculates the first and second order partial derivatives of the concentration of suspended particulates (S.P.) with respect to each of the variables. These are given in Table 2-2 along with measures of sensitivity, the measure of variation in S.P. with respect to that variable over the expected range of variation. Thus, for example, variable X(6) explains the largest variation in S.P. as sensitivity is defined herein. Very briefly, the linear sensitivity is defined as the ratio of the expected change in S.P. to the mean value of S.P. for the expected change in X(i), i = 1, ..., 13. Hence, it is the product of the first partial evaluated at the nominal or mean values of each of the variables multiplied by the expected change in X (say, 2DX) and divided by the nominal value of S.P. This sensitivity measure corresponds to the first order terms in a Taylor series expansion of the model, S.P.=f(X1, ..., X13). The non-linear sensitivity is the contribution of the second order terms in the Taylor series, omitting the mixed terms.

A step-by-step explanation of the procedure used in the sensitivity analysis for estimating the mean and standard deviation using a simple model containing only 3 variables is presented here to help clarify the above discussion. In the modeling of S.P. as a function of the several variables, an approximate linear model is obtained by numerically differentiating the function

$$S.P. = f(X(1), X(2), \ldots, X(13)),$$

obtaining the partial derivatives $\frac{\partial S.P.}{\partial X(1)}, \frac{\partial S.P.}{\partial X(2)}, \ldots, \frac{\partial S.P.}{\partial X(13)}$, and then using a Taylor series approximation with only the first order partial derivatives, thus

$$S.P. \approx \overline{S.P.} + \left.\frac{\partial S.P.}{\partial X(1)}\right|_{\bar{X}} DX(1) + \left.\frac{\partial S.P.}{\partial X(2)}\right|_{\bar{X}} DX(2) + \ldots + \left.\frac{\partial S.P.}{\partial X(13)}\right|_{\bar{X}} DX(13) + \ldots .$$

where $\overline{S.P.}$ is the average value of S.P. calculated using the nominal values of all the variables in the performance model, and

$\left.\frac{\partial S.P.}{\partial X(i)}\right|_{\bar{X}}$ is the value of the i^{th} partial derivative with each variable at its mean value.

From this equation, the mean and variance of S.P. are obtained in the computerized approach as follows. Suppose that for example and simplicity of computation, the following approximation adequately describes the performance model

$$S.P. = 200 + 2DX(1) + 5DX(2) - 3DX(3) \, , \qquad \text{Eq(2-1)}$$

and that the means and variances of the variables are as follows:

	Mean	Variances
DX(1)	0	$4^2 = 16$
DX(2)	0	$5^2 = 25$
DX(3)	3	$2^2 = 4$

Then by substituting the mean values of the variables into Equation 2-1 and performing the calculation, an estimate of 191 is obtained for the mean by

$$\text{Mean (S.P.)} \approx 200 + 2(0) + 5(0) - 3(3) = 191 \, .$$

To estimate the variance of S.P., multiply the square of the coefficients in Equation 2-1 by the respective variances as

$$\text{Variance (S.P.)} = 0 + 2^2\{DX(1)\}^2 + 5^2\{DX(2)\}^2 + 3^2\{DX(3)\}^2$$

and by inserting the standard deviations of the variables and performing the calculations,

$$\text{Variance (S.P.)} = 0 + 2^2(4^2) + 5^2(5^2) + 3^2(2^2) ,$$

that is, the sum of squares of the coefficients by the variances,

$$\text{Variance (S.P.)} = 64 + 625 + 36 = 725$$

or standard deviation of S.P. = $\sqrt{725} = 27$. Note how one variable plays a dominant role, in this case X(2). Typically only a few variables significantly affect the variation of the function. The above computation assumes that the variables are independent; i.e., variation of one variable toward a high (or low) value does not alter knowledge concerning the variation of another variable. If the variables are correlated, an estimate of the variance of S.P. can still be made, given the degree of correlation.

Caution should be observed when using this type approximation. Some of the coefficients of the standard deviations change as the true concentration changes. Until more information is available, it is advisable to derive a new approximation through a computer run any time the concentration level of interest differs more than about ± 20 percent from the concentration used to derive the original approximation. Also, large changes (e.g., ± 20 percent) in the standard deviation will cause a change in the coefficient if the relationship between that variable and the average concentration is nonlinear.

Returning to the analysis discussion using all thirteen variables, the nominal value of S.P. and the standard deviation of S.P. computed by using the first order terms in the Taylor series and performing a usual "error analysis" are given at the bottom of Table 2-2. Referring to Table 2-1 again, there are several additional and useful computations. For example, the "worst case limits" are defined as the largest and smallest values of S.P. using the appropriate combination of the largest and smallest values of each of the variables. Each limit of X(i) is defined by X(i) + 2DX(i), i = 1, 2, ..., 13. The value of X(i) ± 2DX(i) for which S.P. is larger is used in computing the upper (lower) limit. The worst case limits and the nominal value are given below this table of limit values.

The next few lines of the table give a comparison of the "worst case" values of S.P. computed from the model versus those computed from the Taylor series using (1) the first and second degree terms and (2) the first degree terms only. The bottom part of the table contains a check of the individual terms in the Taylor series. The columns headed by S.P.(X - 2DX)/S.P.(X) and S.P.(X + 2DX)/S.P.(X) are the values of S.P. at the respective values X(i) ± 2DX(i) for i = 1, ..., 13, divided by the nominal value of S.P.

The columns headed by 1 ± SENS, 1 ± SENS + NON LIN are the predicted values using the first, and first and second degree terms, respectively, of the series. Note the good agreement throughout this tabulation.

The significant feature of this analysis is that it provides a measure of where the significant variations in the S.P.'s result from the analysis or measurement process. In this example, it is clear that (see Table 2-2) X(6), X(7), and X(4) are the three largest contributing variables to the variation in S.P. Hence, quality control and assurance techniques which will minimize these variations; namely, errors in flow-rate calibration, variations resulting from changes in temperature and/or pressure, and errors in flow-rate readings should be looked at first. In this particular run, it is assumed that the pH of the filter and the relative humidity of the conditioning environment, RH, are controlled within narrow limits. Further uses of the results of the sensitivity analysis in the decision process are discussed in the following paragraphs.

The previous steps in the analysis have identified the important variables in the measurement process and have estimated to what degree they affect the value of S.P. The sensitivity analysis also provides an approximate model of the process which may be written directly from the printout (Table 2-2) as

$$\begin{aligned} \text{S.P.} = {} & 100 + 0.531\,DX(1) + 0.133\,DX(2) + 466.8\,DX(3) \\ & - 98.2\,DX(4) - 47.1\,DX(5) - 98.2\,DX(6) - 98.2\,DX(7) \\ & - 98.2\,DX(8) + 48.0\,DX(9) - 0.106\,DX(10) - 0.0583\,DX(11) \\ & - 0.0895\,DX(12) + 3.38\,DX(13). \end{aligned}$$

The coefficients of the DX(i), i = 1, ..., 13 are the first order partial derivatives read directly from Table 2-2 under S.P.'. This model approximation may be used to relate the variation in S.P. as estimated by the standard deviation in S.P. to given variations or errors in the X(i)'s. The adequacy of this approximation is checked thoroughly in the sensitivity analysis. By introducing more control or improved instrumentation, one can reduce the standard deviation in the X(i)'s, and hence, the standard deviation of S.P. The relative cost of providing the additional control or instrumentation must be weighed against the improved precision and/or accuracy of the results. The computer program can be used in the decision process as follows: Change the variation DX(i) of the particular X(i) of interest to the value expected with improved instrumentation or introduction of a quality control procedure, and compute the estimated standard deviation in S.P. as was illustrated previously using a simplified model.

Coefficients in the above approximation were obtained for a true average concentration of 100 $\mu g/m^3$ (approximated by the constant term) and with the variables modeled as shown in Table 2-1, page 133, the two columns titled "Mean" and "Deviation". A new approximation must be generated by a computer run for other concentration levels or when a variable model is significantly changed.

This approach gives an objective procedure for making decisions relative to equipment needs, need for further data on critical variables, and where to introduce more quality control and assurance, sampling, etc., in order to obtain the desired improvements in the results. The day-to-day variation in concentration of particulate matter are not included in the analysis given by the model.

The decision model derived in the sensitivity analysis is subject to the validity and adequacy of available data and statistical and engineering judgments when data are not available. If the model is based on too many assumptions without any supporting data, it is desirable to plan an experimental program to estimate the variation in S.P. resulting from variations or errors in the measurement process. The sensitivity analysis should be used to identify the important variables; interactions, if any, between the variables; and the expected effect of the variables on S.P. This information will be useful in designing an efficient experimental program to improve the model. Such an approach should have most value in the initial stages of evaluating new instruments and/or analytical procedures.

TABLE 2-1: Partial Output of Sensitivity Analysis

```
WORST CASE LIMITS                                                        Mean            Deviation
VALUE OF VARIABLE AT LOWER LIMIT  AND AT UPPER LIMIT ,                     X                 DX
        X1                .65000E  1          .12500E  2           .95000E  1        .15000E  1
        X2                .45000E  2          .65000E  2           .55000E  2        .50000E  1
        X3               -.20000E -2          .20000E -2           .00000E  0        .10000E -2
        X4                .50000E -1         -.50000E -1           .00000E  0        .25000E -1
        X5                .10000E  0         -.20000E -1           .40000E -1        .30000E -1
        X6                .10000E  0         -.10000E  0           .00000E  0        .50000E -1
        X7                .60000E -1         -.60000E -1           .00000E  0        .30000E -1
        X8                .40000E -1         -.40000E -1           .00000E  0        .20000E -1
        X9               -.70000E -1          .70000E -1           .00000E  0        .35000E -1
        X10               .21000E  2          .10000E  1           .11000E  2        .50000E  1
        X11               .30000E  2          .10000E  2           .20000E  2        .50000E  1
        X12               .10000E  2         -.10000E  2           .00000E  0        .50000E  1
        X13               .80000E  0          .16400E  1           .12200E  1        .21000E  0

WORST CASE LIMITS AND NOMINAL VALUE
      SP                  .64950E  2          .14476E  3           .10007E  3

INTERACTION CHECK USING 1ST AND 2ND DEGREE TERMS OF TAYLOR SERIES
      SP                  .64142E  2          .14083E  3
INTERACTION CHECK USING 1ST DEGREE TERMS OF TAYLOR SERIES
      SP                  .61724E  2          .13841E  3

GOODNESS OF FIT USING 1ST AND 2ND TERMS OF TAYLOR SERIES
VARIABLES  SP(X-2DX)/SP(X)    1.-SENS    1.-SENS+NON LIN   SP(X+2DX)/SP(X)    1.+SENS    1.+SENS+NON LIN
    X1      .98407E  0     .98407E  0     .98406E  0        .10159E  1     .10159E  1     .10159E  1
    X2      .99113E  0     .98674E  0     .99251E  0        .10212E  1     .10133E  1     .10190E  1
    X3      .99067E  0     .99067E  0     .99067E  0        .10093E  1     .10093E  1     .10093E  1
    X4      .10516E  1     .10490E  1     .10515E  1        .95324E  0     .95095E  0     .95335E  0
    X5      .10291E  1     .10283E  1     .10290E  1        .97252E  0     .97174E  0     .97254E  0
    X6      .11088E  1     .10981E  1     .11077E  1        .91065E  0     .90189E  0     .91151E  0
    X7      .10625E  1     .10589E  1     .10623E  1        .94440E  0     .94113E  0     .94460E  0
    X8      .10408E  1     .10392E  1     .10408E  1        .96223E  0     .96076E  0     .96230E  0
    X9      .96642E  0     .96642E  0     .96641E  0        .10336E  1     .10336E  1     .10336E  1
    X10     .10106E  1     .10106E  1     .10106E  1        .98941E  0     .98941E  0     .98940E  0
    X11     .10058E  1     .10058E  1     .10058E  1        .99417E  0     .99417E  0     .99417E  0
    X12     .10070E  1     .10069E  1     .10070E  1        .99310E  0     .99305E  0     .99309E  0
    X13     .98641E  0     .98580E  0     .98635E  0        .10147E  1     .10142E  1     .10147E  1
```

TABLE 2-2: Printout of Sensitivity Analysis of S.P.

```
FIRST AND SECOND PARTIAL DERIVATIVES (S.P.' and S.P.") OF S.P. WITH RESPECT TO X
                                                                   PARTIALS                SENSITIVITY             Ranking of Linea
 X     S.P.(X-2DX)  S.P.(X-1DX) S.P.(X+1DX)  S.P.(X+2DX)  S.P.'          S.P."          LINEAR      NON-LIN          Sensitivities

 X1   .98477E  2  .99274E  2  .10087E  3  .10167E  3  .53134E  0 -.13563E -3  .15929E -1 -.60991E -5
 Y2   .99184E  2  .99533E  2  .10090E  3  .10219E  3  .13264E  0  .11535E -1  .13254E -1  .57632E -2
 X3   .99138E  2  .99604E  2  .10054E  3  .10100E  3  .46684E  3 -.20345E  2  .93300E -2 -.40661E -6
 X4   .10523E  3  .10259E  3  .97676E  2  .95392E  2 -.98176E  2  .19242E  3 -.49053E -1  .24035E -2
 X5   .10298E  3  .10151E  3  .98677E  2  .97321E  2 -.47125E  2  .44510E  2 -.28255E -1  .80061E -3
 X6   .11096E  3  .10523E  3  .95392E  2  .91130E  2 -.98181E  2  .19255E  3 -.98110E -1  .96208E -2
      .10633E  3  .10311E  3  .97210E  2  .94508E  2 -.98186E  2  .19271E  3 -.58869E -1  .34664E -2
 X8   .10416E  3  .10287E  3  .98146E     .96292E  2 -.98178E  2  .19267E  3 -.39243E -1  .15402E -2
 X9   .96711E  2  .98391E  2  .10175E  3  .10343E  3  .48002E  2 -.28234E  0  .33577E -1 -.69124E -5
 X10  .10113E  3  .10060E  3  .99541E  2  .99012E  2 -.10599E  0 -.13021E -4 -.10599E -1 -.65058E -5
 X11  .10065E  3  .10036E  3  .99780E  2  .99488E  2 -.58305E -1  .81380E -6 -.58263E -2  .40661E -6
 X12  .10077E  3  .10042E  3  .99725E  2  .99381E  2 -.69515E -1  .83821E -4 -.69465E -2  .41881E -4
 X13  .98712E  2  .99375E  2  .10079E  3  .10154E  3  .33828E  1  .61866E  0  .14198E -1  .54526E -3
ALL X AT NOMINAL, S.P.(X) =.10007E  3

STD DEV OF S.P.(X), .70540E  1
```

SENSITIVITY ANALYSIS OF THE PARAROSANILINE METHOD OF MEASURING SO_2

The same type of sensitivity analysis as was described in the previous section of this Appendix for the High Volume Method was performed for the Pararosaniline Method of measuring SO_2.

The input values of the mean (X) and standard deviation (DX) of each of the fifteen variables are given in the last two columns of the upper part of Table 2-3. These are the same values used in Table 5 of Section II of this document.

The column titled Y@ (Y@ represents the first derivative) in Table 2-4 contains the coefficients of the linear terms of the Taylor series approximation of the performance model at a true SO_2 concentration of 800 μg SO_2/m^3. This represents computer run number 1 given in Table 7 of Section II. The variables can be ranked according to their effect on the measured concentration by ordering the values given in the next to last column of Table 2-4. In this case only linear terms are important since all non-linear (second degree) terms are zero (last column in the table). An approximate model for estimating the average measured value composed of the five most important variables is

$$Y = 757.6-9.59\ DX(13) + 75.76\ DX(5) + 25254\ DX(1)-1156DX(14) + 789\ DX(6).$$

To see how much of the total variance is accounted for by these five variables, substitute their individual standard deviations, as given in Table 2-3, in the following relationship

$$\sigma^2(Y) = 0+(9.59)^2(5)^2+(75.76)^2(.28)^2 + (25254)^2(.00082)^2 + (1156)^2(.015)^2$$
$$+(789)^2(.02)^2,$$

$$\sigma^2(Y) = 4118,$$

$$\sigma(Y) = 64.$$

Then 64 is approximately 95 percent of 67.44 which is the standard deviation given in the last line of Table 2-4 and was computed using all 15 variables. In this instance then quality control procedures should be directed toward these 5 variables, specifically X13 which accounts for approximately 50 percent of the total variability.

The above approximate model is only valid for SO_2 concentrations around 800 μg SO_2/m^3. Approximate models for concentrations of 400 μg SO_2/m^3 and 100 μg SO_2/m^3 were obtained from computer runs 2 and 3, respectively, as given in Table 7 of Section II.

For SO_2 concentrations close to 400 μg SO_2/m^3 the following model containing the five most influential variables is

$$Y = 379-4.8\ DX(13)-1156\ DX(4) + 1156\ DX(3)+38DX(5)+12627\ DX(1).$$

A similar approximate model for concentrations around 100 μg SO_2/m^3 is

$$Y = 95-1156\ DX(4)+1156\ DX(3) - 1.2\ DX(13)+9.5\ DX(5)+3157\ DX(1).$$

The changes in the coefficients for the above models result from the interactions in the original model; that is, it contains several product terms of the form X(1) X(2).

TABLE 2-3: Partial Output of Sensitivity Analysis (SO_2)

```
  WORST CASE LIMITS                                                              X                    DX
  VALUE OF VARIABLE AT LOWER LIMIT  AND AT UPPER LIMIT .
       X1              .28360E -1              .31640E -1              .30000E -1          .82000F -3
       X2              .16300E  0              .16300E  0              .16300E  0          .12000E -1
       X3             -.24000E -1              .24000E -1              .00000E  0          .12000E -1
       X4              .30000E -1             -.30000E -1              .00000E  0          .15000E -1
       X5              .94400E  1              .10560E  2              .10000E  2          .29000E  0
       X6             -.40000E -1              .40000E -1              .00000E  0          .20000F -1
       X7             -.40000E -1              .00000E  0             -.20000E -1          .10000E -1
       X8             -.40000E -1              .00000E  0             -.20000E -1          .10000E -1
       X9             -.22400E -1              .22400E -1              .00000E  0          .11200E -1
       X10            -.11400E -1              .11400E -1              .00000E  0          .57000F -2
       X11            -.14000E  2              .14000E  2              .00000E  0          .70000F  1
       X12            -.50000E -1              .00000E  0             -.25000E -1          .12500F -1
       X13             .25000E  2              .50000E  1              .15000E  2          .50000E  1
       X14             .00000E  0              .10000E  2              .50000E  1          .25000F  1
       X15            -.50000E -1              .00000E  0             -.25000E -1          .12500E -1

  WORST CASE LIMITS AND NOMINAL VALUE
     SEQ1              .47892E  3              .13446E  4              .75762E  3

  INTERACTION CHECK USING 1ST AND 2ND DEGREE TERMS OF TAYLOR SERIES
     SEQ1              .35113E  3              .11641E  4
  INTERACTION CHECK USING 1ST DEGREE TERMS OF TAYLOR SERIES
     SFQ1              .35113E  3              .11641E  4

  GOODNESS OF FIT USING 1ST AND 2ND TERMS OF TAYLOR SERIES
  VARIABLES   Y(X-2DX)/Y(X)    1.-SFNS     1.-SENS+NON LIN   Y(X+2DX)/Y(X)     1.+SENS      1.+SFNS+NON LIN
      X1      .94534F  0    .94533E  0    .94533E  0      .10547F  1     .10547F  1      .10547E  1
      X2      .10000F  1    .10000E  1    .10000E  1      .10000E  1     .10000F  1      .10000E  1
      X3      .96337F  0    .96337E  0    .96337E  0      .10366E  1     .10366F  1      .10366E  1
      X4      .10458F  1    .10458E  1    .10458E  1      .95421E  0     .95422F  0      .95422E  0
      X5      .94400F  0    .94400E  0    .94400E  0      .10560F  1     .10560F  1      .10560E  1
      X6      .95833F  0    .95833E  0    .95833E  0      .10417E  1     .10417E  1      .10417E  1
      X7      .97917F  0    .97917E  0    .97917E  0      .10208F  1     .10208F  1      .10208E  1
      X8      .97917F  0    .97917E  0    .97917E  0      .10208F  1     .10208F  1      .10208E  1
      X9      .97667F  0    .97666E  0    .97666E  0      .10233E  1     .10233F  1      .10233E  1
      X10     .98813F  0    .98812E  0    .98812E  0      .10119E  1     .10119F  1      .10119E  1
      X11     .99028F  0    .99027E  0    .99027E  0      .10097E  1     .10097F  1      .10097E  1
      X12     .97468F  0    .97468E  0    .97468E  0      .10253E  1     .10253F  1      .10253E  1
      X13     .11266F  1    .11266E  1    .11266E  1      .87341E  0     .87342F  0      .87342E  0
      X14     .96203F  0    .96203E  0    .96203E  0      .10380E  1     .10380F  1      .10380E  1
      X15     .97468F  0    .97468E  0    .97468E  0      .10253E  1     .10253F  1      .10253E  1
```

TABLE 2-4: Printout of Sensitivity Analysis (SO_2)

```
FIRST AND SECOND PARTIAL DERIVATIVES (Y* AND Y**) OF SEQ1 WITH RESPECT TO Y
                                                          PARTIALS                  SENSITIVITY
 X -  Y(X-2DX)     Y(X-1DX)     Y(X+1DX)     Y(X+2DX)     Y*           Y**          LINEAR       NON-LIN
 X1   .71621E  3  .73691E  3  .77833E  3  .79904E  3  .25254E  5  .00000E  0  .54666E -1  .00000E  0
      .20708E  2  .87000E -3  .00000E  0  .67240E -6
 X2   .75762E  3  .75762E  3  .75762E  3  .75762E  3  .00000E  0  .00000E  0  .00000E  0  .00000E  0
      .00000E  0  .12000E -1  .00000E  0  .14400E -3
 X3   .72987E  3  .74374E  3  .77150E  3  .78537E  3  .11364E  4  .00000E  0  .36631E -1  .00000E  0
      .13876E  2  .12000E -1  .00000E  0  .14400E -3
 X4   .79231E  3  .77496E  3  .74028E  3  .72293E  3 -.11562E  4  .00000E  0 -.45784E -1  .00000E  0
     -.17344E  2  .15000E -1  .00000E  0  .22500E -3
 X5   .71519E  3  .73641E  3  .77883E  3  .80005E  3  .75760E  2  .00000E  0  .55998E -1  .00000E  0
      .21213E  2  .28000E  0  .00000E  0  .78400E -1
 X6   .72606E  3  .74184E  3  .77341E  3  .78919E  3  .78922E  3  .00000E  0  .41668E -1  .00000E  0
      .15764E  2  .20000E -1  .00000E  0  .40000E -3
 X7   .74184E  3  .74973E  3  .76551E  3  .77341E  3  .78888E  3  .00000E  0  .20825E -1  .00000E  0
      .78888E  1  .10000E -1  .00000E  0  .10000E -3
 X8   .74184E  3  .74973E  3  .76551E  3  .77341E  3  .78888E  3  .00000E  0  .20825E -1  .00000E  0
      .78888E  1  .10000E -1  .00000E  0  .10000E -3
 X9   .73995E  3  .74878E  3  .76646E  3  .77530E  3  .78927E  3  .00000E  0  .23336E -1  .00000E  0
      .88398E  1  .11200E -1  .00000E  0  .12544E -3
 X10  .74863E  3  .75312E  3  .76212E  3  .76662E  3  .78936E  3  .00000E  0  .11877E -1  .00000E  0
      .44993E  1  .57000E -2  .00000E  0  .32490E -4
 X11  .75026E  3  .75394E  3  .76131E  3  .76499E  3  .52625E  0  .00000E  0  .97244E -2  .00000E  0
      .36838E  1  .70000E  1  .00000E  0  .49000E  2
 X12  .73844E  3  .74803E  3  .76722E  3  .77680E  3  .76738E  3  .00000E  0  .25322E -1  .00000E  0
      .95923E  1  .12500E -1  .00000E  0  .15625E -3
 X13  .85353E  3  .80557E  3  .70967E  3  .66172E  3 -.95899E  1  .00000E  0 -.12658E  0  .00000E  0
     -.47950E  2  .50000E  1  .00000E  0  .25000E  2
 X14  .72886E  3  .74324E  3  .77201E  3  .78639E  3  .57534E  1  .00000E  0  .37970E -1  .00000E  0
      .14384E  2  .25000E  1  .00000E  0  .62500E  1
 X15  .73844E  3  .74803E  3  .76722E  3  .77680E  3  .76738E  3  .00000E  0  .25322E -1  .00000E  0
      .95923E  1  .12500E -1  .00000E  0  .15625E -3
ALL X AT NOMINAL. Y(X) =   .75762E  3
STD DEV OF Y(X),    .67437E  2
```

APPENDIX 3 FREQUENCY DISTRIBUTIONS

NORMAL (GAUSSIAN) FREQUENCY DISTRIBUTION

A brief discussion concerning the normal or gaussian frequency distribution is given herein. The normal frequency distribution has been used frequently throughout the field documents and in this final report to describe the variation of a conceptual population of measurements. For example, suppose that measurements are being made of the concentration of a specific pollutant in a control gas sample. Conceptually a very large number of such measurements could be made, generating a frequency distribution of these values. These data will have an average or mean value, the true value of the concentration, and a dispersion about this value measured by the root mean square of the deviations from the mean. The latter statistic (corrected for bias in the variance estimate) is referred to as the standard deviation. The mean and the standard deviation completely specify the normal distribution. That is, from the use of these parameters one can answer any question concerning the likelihood of a measurement taken at random falling within any specific interval. (Notation used throughout this document to represent a normal distribution with a mean value μ and a standard deviation σ is $N(\mu,\sigma)$.)

If the mean and standard deviation are denoted by μ and σ, respectively, the normal frequency distribution is given by the mathematical formula

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

where x is the measurement, exp {y} denotes e^y, where e = 2.7128. This frequency distribution is bell-shaped, symmetrical with the mean as the point of symmetry. Figure 3-1 is a sketch of the curve.

In order to apply this frequency distribution to a specific problem, it is first necessary to decide that the distribution is an adequate representation of the data. This can be decided on the basis of data, if available, or by intuition and the fact that so many measurements are adequately approximated by such a frequency distribution. It is next necessary to estimate the mean and standard deviation of the measurements. With these values one can estimate the relative frequency of measurements in any specified interval, say (a,b), by obtaining the proportion of the total area under the curve which falls between the lines x=a and x=b; see Figure 3-1. These areas are tabulated in tables for areas under the normal curve. For example, in Reference 27, one obtains the area to the left of b and subtracts from this the area to the left of a. To enter the table, one must compute standard normal variables by computing

$$\frac{b-\mu}{\sigma} \text{ and } \frac{a-\mu}{\sigma},$$

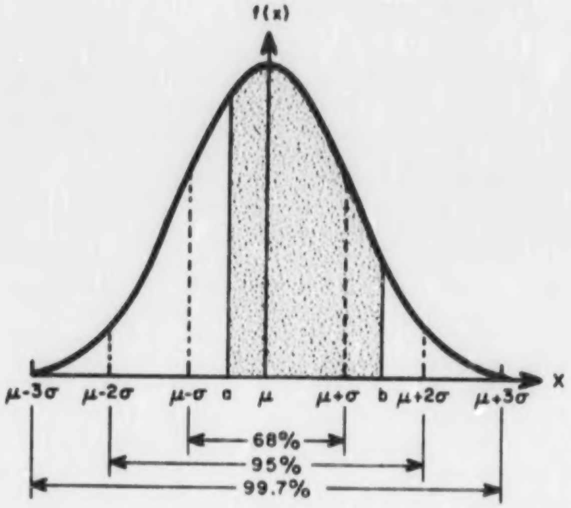


Figure 3-1: Sketch of Normal Frequency Distribution

and then, looking up the corresponding probability (area) that x is less than b, a, respectively,

$$\text{Prob}(a < x \le b) = \text{Prob}(x \le b) - \text{Prob}(x \le a) .$$

For example, if $b = 25$, $a = 10$, $\mu = 15$, $\sigma = 5$, we have

$$P(10 < x \le 25) = P(x \le 25) - P(x \le 10) .$$

Converting the values 25 and 10 to standard normal values u, one obtains

$$= P\left(u \le \frac{25-15}{5}\right) - P\left(u \le \frac{10-15}{5}\right)$$

$$P(10 \le x \le 25) = 0.97725 - .1587 \approx 0.82$$

A few useful values are given below for convenient reference.

Standard Normal Variable $\left(u = \frac{x-\mu}{\sigma}\right)$	Probability that x falls between - u and + u
1.0	0.68
1.645	0.90
1.96	0.95
2.58	0.99
3.0	0.997

For example, the probability that a measurement falls between $\mu \pm 2\sigma$ ($\pm 3\sigma$) is approximately 0.95 (0.997).

In some cases there is very definite evidence that the data are not normally distributed. For example, the concentration of particles in ambient air is very adequately approximated by the logarithmic normal frequency distribution. That is, the logarithm of the concentration is normally distributed. Thus, one cannot take the normal frequency distribution on complete faith without checking its adequacy. There are statistical tests for making such an adequacy check. (Plotting the data on normal probability graph paper is a simple visual check.)

There is one fundamental theorem which gives considerable justification for assuming the normal frequency distribution. This is referred to as the central limit theorem. It indicates that the average or mean of a large number of measurements tends to be normally distributed even when an individual measurement may not be very well approximated by this same distribution. Further discussion of the implications of this theorem and of the normal frequency function is given in many standard texts; e.g., see Ref 26.

The normal distribution has another important characteristic: a linear combination of normally distributed variables is also normally distributed. See the section on Means and Variances of Linear Combinations of Variables for a more detailed discussion of how one can obtain or compute the mean and variance of the linear combination. This result is used frequently in explaining the overall variation of results. For example, the measurements of the concentration of a control gas sample by several monitoring stations (laboratories) might be considered as a sum of the variation among laboratories and that within laboratories, say,

$$C_{ij} = \mu + \ell_i + w_j$$

where C_{ij} = j^{th} measured concentration of the sample by the i^{th} lab,

μ = the overall (true) mean concentration,

ℓ_i = deviation of the mean concentration for laboratory i from the overall mean,

w_j = deviation of the j^{th} individual (within) laboratory measurement from the lab mean.

Now if ℓ_i and w_j are assumed to be normally distributed, C_{ij} is also normally distributed with mean μ and variance $\sigma_\ell^2 + \sigma_w^2$, the sum of the variances of the ℓ's and the w's. The ℓ_i might also be considered as a laboratory bias; the above model assumes these biases have mean zero and standard deviation σ_ℓ. Such a simple model may also be generalized to include operator effects, day effects, instrument errors or variation, etc.

UNIFORM FREQUENCY DISTRIBUTION

The uniform distribution with the parameters a and b is defined by the equation

$$f(x) = \begin{cases} 1/(b-a) & \text{for } a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

and its graph is shown in Figure 3-2. (Notation used throughout this document to represent a uniform distribution with a lower limit a and an upper limit of b is U(a to b).) All values of x from a to b are "equally likely" in the sense that the probability that x lies in a narrow interval of width Δx entirely contained in the interval from a to b is equal to $\Delta x/(b-a)$, regardless of the exact location of the interval.

The governing equations for a uniform distribution are:

$$(1) \quad f(x) = 1/(b-a)$$

$$(2) \quad \mu = (a+b)/2, \text{ and}$$

$$(3) \quad \sigma = (b-a)/\sqrt{12} \ .$$

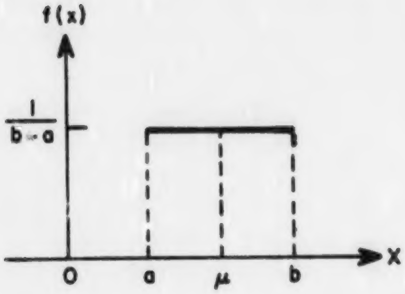


Figure 3-2: Sketch of Uniform Frequency Distribution

APPENDIX 4 ESTIMATION OF THE MEAN AND VARIANCE OF VARIOUS COMPARISONS OF INTEREST

MEAN AND VARIANCE OF LINEAR FUNCTIONS

Difference of Two Observations

In the development of standards for comparison of measured values with audited values and of measured values with the predicted values given by a calibration, it is necessary to estimate the variance of simple linear combinations of observations. For example, in comparing a measured value with an audited value such as in a data processing check, the variance of the difference is

$$\sigma^2(X_1 - X_2) = \sigma_1^2 + \sigma_2^2,$$

where σ_1^2 is the variance of the first measured value, and

σ_2^2 is the variance of the second measured value, the audited value.

In practice these two variances may be assumed to be equal to the same value, σ_1 say, particularly if the measurements are made by two different operators using the same instrument. Thus under this assumption,

$$\sigma^2(X_1 - X_2) = 2\sigma_1^2,$$

and if one obtains the estimated variance of the differences of n pairs of measurements, the result would estimate $2\sigma_1^2$ rather than σ_1^2, or

$$\frac{s^2(X_1 - X_2)}{2}$$

would estimate σ_1^2. This is the reason for inserting a 2 in the denominator of some of the expressions for s, the computed standard deviation, in the field documents.

The mean of the difference of two observations is the difference of their means, i.e., if $\mu(X)$ denotes the mean or expected value of X, then

$$\mu(X_1 - X_2) = \mu(X_1) - \mu(X_2) \text{ or } \mu_1 - \mu_2.$$

It is expected that μ_1 would equal μ_2 if both measurements are of the same characteristic such as the final filter weight, unless an operator is biased, e.g., tends to read a scale too high for all measurements. If $\mu_1 = \mu_2$ then $\mu(X_1 - X_2) = 0$ by assumption. This assumption can be checked by a standard statistical test, called the t test (Ref. 26).

Suppose that a comparison is being made of a measured value and an audited value which is considered to be very precise, e.g., a case in which a very precise instrument or "known" control sample (i.e., with a precisely determined concentration) is being used in the audit check. In this case the variance of the difference will almost be equivalent to the variance of the measured value with the larger variance σ_1^2, say, because the addition of σ_2^2 to σ_1^2 will alter the total result very little. For example, if $\sigma_1 = 5$, $\sigma_2 = 1$, the $\sigma_1^2 + \sigma_2^2 = 26$ compared to $\sigma_1^2 = 25$, assuming the measurements are uncorrelated. Furthermore, $\sigma_1 = 5$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} = 5.2$. In some cases in the field documents it is assumed that one is comparing a measured value with a relatively precise value, and the estimated variance desired for computing the precision of the measurement process is that of the measured value. Hence, the assumption is made that the variance of the difference is essentially equal to the variance of the less precise measured value. The mean difference will be zero unless some bias is present in the measurement method, e.g., the interference from water vapor in the CO measurement.

Mean of n Observations

The most frequent application of estimating the variance of a linear combination is that of the mean of n observations. For n=2,

$$\sigma^2\left(\frac{X_1 + X_2}{2}\right) = \frac{\sigma^2}{4} + \frac{\sigma^2}{4} = \frac{\sigma^2}{2}$$

or in general

$$\sigma^2\left(\frac{X_1 + X_2 + \ldots + X_2}{n}\right) = \sigma^2(\bar{X}) = \frac{\sigma^2}{n} .$$

This is a well-known result of the standard error for the mean, i.e.,

$$\sigma(\bar{X}) = \sigma/\sqrt{n} .$$

The mean of the average of n measurements is the overall mean, or "population" mean, if there is no measurement bias, i.e., $\mu(\bar{X}) = \mu$. This application is used in the field documents in the case of the data processing check, where averages of two differences are reported if required by the manager Each difference has a variance of $2\sigma^2$, and hence the average of two differences has the variance $2\sigma^2/2 = \sigma^2$, the same variance as that of an individual measurement. This fact is used in the data processing check in the field documents.

General Case

In general, if Y is a linear combination of n values X_1, X_2, ..., X_n, such as

$$Y = \ell_o + \ell_1 X_1 + \ell_2 X_2 + \ldots + \ell_n X_n ,$$

then if the X's are independent or uncorrelated, the variance of Y is given by

$$(1) \qquad \sigma^2(Y) = \ell_1^2\sigma_1^2 + \ell_2^2\sigma_2^2 + \ldots + \ell_n^2\sigma_n^2 ,$$

and if $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$,

$$(1a) \qquad \sigma^2(Y) = \sigma^2 \times \sum_{i=1}^{n} \ell_i^2 .$$

The mean of the linear combination is the same linear combination of the means, i.e.,

$$(2) \qquad \mu_1(Y) = \ell_o + \ell_1\mu_1 + \ell_2\mu_2 + \ldots + \ell_n\mu_n = \ell_o + (\Sigma\ell_i)\mu$$

if all $\mu_i = \mu$, $i = 1, 2, \ldots, n$. If the X_i are not uncorrelated but have correlation coefficients ρ_{ij} for X_i, X_j, the variance of Y becomes

$$(3) \qquad \sigma^2(Y) = \ell_1^2\sigma_1^2 + \ell_2^2\sigma_2^2 + \ldots + \ell_n^2\sigma_n^2$$

$$+ 2\ell_1\ell_2\rho_{12}\sigma_1\sigma_2 + 2\ell_1\ell_3\rho_{13}\sigma_1\sigma_3 + \ldots + 2\ell_{n-1}\ell_n\rho_{n-1,n}\,\sigma_{n-1}\sigma_n .$$

If the σ_i's are all equal, then

$$(3a) \qquad \sigma^2(Y) = \sigma^2(\Sigma \ell_i^2) + 2\sigma^2 \, \Sigma_{(i<j)} \Sigma \ell_i \ell_j \, \rho_{ij} \; .$$

For example, suppose that several measurements are obtained on the concentration of suspended particles ($\mu g/m^3$) at neighboring stations and that it is known that they are correlated with $\rho = 0.5$. The variance of the difference of two such measurements would be

$$\sigma^2(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 - 2\rho \; \sigma_1 \sigma_2 .$$

If $\sigma_1 = \sigma_2$, then for $\rho = 0.5$,

$$\sigma^2(X_1 - X_2) = 2\sigma^2 - 2\left(\frac{1}{2}\right)\sigma^2$$

$$= \sigma^2$$

and not $2\sigma^2$ as it would be for uncorrelated measurements.

Comparison with Calibrated Value

Suppose that a calibration is made at 5 levels. A procedure for comparing a single observation with a calibrated value given by a straight line relationship, using tolerance intervals, is given in Ref. 28 and presented briefly here.

Consider an example in which an analyzer is calibrated using a series of five samples of "known" concentrations distributed over the range of interest. Let y be the analyzer reading in volts and x the concentration of the sample in $\mu g/m^3$. Assume that

$$y = \beta_o + \beta_1 \; (x_i - \bar{x}) + \varepsilon ,$$

that is, the observed analyzer reading is a linear function of x with unknown coefficients to be estimated from the data and ε is the deviation of the observed analyzer reading from the true value given by $\beta_o + \beta_1(x_i - \bar{x})$.

In this example it is assumed that x is without error. The unknown coefficients β_o and β_1 can be estimated by the method of least squares. Let their estimates be b_o and b_1, respectively, then

$$b_o = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{5} y_i}{5} = \bar{y}$$

and

$$b_1 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})y_i}{\sum_{i=1}^{5} (x_i - \bar{x})^2}$$

Then the predicted mean value of y for given x, say Y, will be given by

$$Y_i = b_o + b_1 (x_i - \bar{x}),$$

and $y_i - Y_i$ is the deviation of the predicted value from the observed value. The b's were chosen so that the sum of squares of these deviations for the n = 5 values would be a minimum. Now s, the estimated standard deviation of the y's for a given x is given by

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - Y_i)^2}{n - 2} \text{ or } \frac{\sum_{i=1}^{5} y_i^2 - b_o \left(\sum_{i=1}^{5} y_i\right) - b_1 \sum_{i=1}^{5} (x_i - \bar{x})y_i}{3}$$

where n = 5. It has been implicitly assumed that the variance of the y's for a given x is independent of x. This is not always the case and, in fact, there is considerable evidence to indicate otherwise in the analysis of pollutant data. In order to consider the case of unequal variances, it is suggested that one refer to Ref. 26, pages 551-7. In this example we will assume constant variance for simplicity of discussion. The modification of the techniques for the heterogeneous variance case is reasonably straightforward.

Now consider the comparison of another determination of y for a given x, say x_o. For example, suppose that a control sample is given to the operator to check his instrument, calibration, etc., and that he estimates the concentration of the sample by the analyzer reading and the calibration curve he is currently using. Let his reading be y_o. Determine if the predicted concentration, x_1, given by the calibration curve and the known concentration are consistent. This is equivalent to requiring that the observation y_o does not deviate too far from the Y_o given by the calibration curve. See Figure 4-1 below. The variance of a single value y_o for a given x,

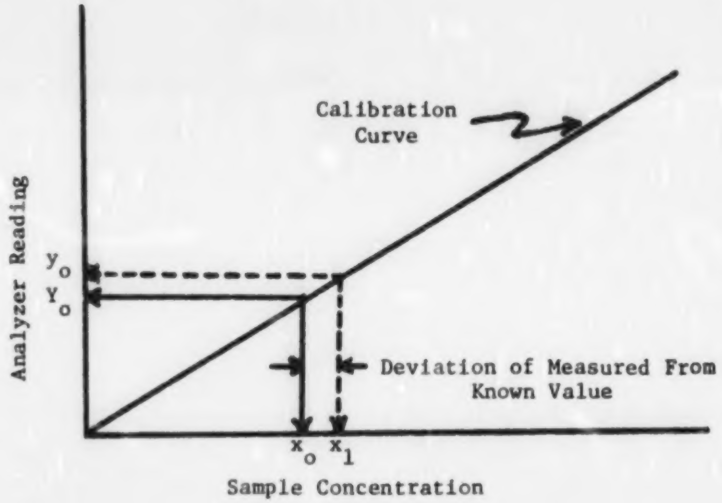


Figure 4-1: Comparison of Individual Analyzer Reading With Value From Calibration Curve

say x_o, is given by σ^2 $(Y_o + e)$, i.e.,

$$\sigma^2 (b_o + b_1 (x_o - \bar{x}) + e) = \sigma^2\left(\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} + 1\right),$$

(e denotes the "error" associated with the individual measurement). This formula is appropriate if one is comparing a single determination of y_o with a single "calibrated" value. However, if one wishes to do this repeatedly, a tolerance interval must be obtained (Ref. 28).

For information concerning the fact that the x's are not without error, see Ref. 22 for discussion of this assumption and some appropriate techniques.

An example computation is given below for illustrating the computations suggested above.

x (concentration in ppm)	y (analyzer reading in volts)
10	.039
20	.086
30	.140
50	.254
75	.369

Computations

$\bar{x} = 37 \qquad \bar{y} = 0.1776$

$b_o = 0.1776, \qquad b_1 = 0.0052$

$Y = 0.1776 + 0.0052\ (x - 37)$

$s^2 = 0.000042$

$s = 0.0065$

$$s^2(y_o) = s^2 \left(1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum_{i=1}^{5}(x_i - \bar{x})^2}\right).$$

Now n=5, and $\sum_{i=1}^{5}(x_i - \bar{x})^2 = 2010$, and let x_o be 30, then

$$s^2(y_o) = 0.000042 \left(1 + \frac{1}{5} + \frac{49}{2010}\right)$$

$$\approx 0.0001 .$$

$$s(y_o) \approx 0.01$$

A 90% confidence interval estimate for the mean of the y's for $x_o = 30$ is given by

$$Y_o \pm 2.353\ s(y_o)$$

where $Y_o = 0.1776 + 0.0052\ (30-37) = 0.1412$ and 2.353 is the t-value (t-distribution) for 90% confidence and for 3 degrees of freedom ($3 = n-2$ in this example) associated with the estimate of s. This procedure is sufficient if we are interested in making just one such statement using one given calibration. However, if we wish to repeat this many times, a tolerance interval is required, as indicated above. The computation of the tolerance limits requires a value of k such that

$$Y_o \pm ks$$

includes, say, 95% of the observations with a desired confidence. Following the procedure in Ref. 28 yields $k \approx 5$ (it ranges from 4.89 at $x=30$ to 5.74 at $x=75$) to include 95% of the observations with 90% confidence. These two intervals are wide due to the small number of observations (5) and the corresponding small number of degrees of freedom. If five calibrations were made and averaged, then $n=25$, the degrees of freedom would be $n-2=23$, and the intervals corresponding to the above would be

Confidence Interval: $Y_o \pm 1.71\ s(y_o)$

and

Tolerance Interval: $Y_o \pm 2.49\ s$, at $x = 30$,

respectively, where $s = 0.006$

$$s^2(y_o) = 0.000042 \left(1 + \frac{1}{25} + \frac{1}{5}\left(\frac{49}{2010}\right)\right) = 0.000044$$

or

$$s(y_o) = 0.0066.$$

Thus, not only $s(y_o)$ has decreased, but the t and k values have decreased. Some of the pertinent computations are given in Table 4-1. The resulting interval for N = 25 is about one-half as wide as that for N = 5.

Figure 4-2 graphically illustrates a calibration curve with confidence limits for a single analyzer reading and tolerance limits for 95% of all analyzer readings.

Table 4-1: COMPUTATION OF CONFIDENCE AND TOLERANCE LIMITS FOR THE EXAMPLE

N = 5 Observations, Single Calibrations

x	$\frac{1}{N'} = \frac{1}{5} + \frac{(x - 37)^2}{2010}$	k	ks	Y	Y - ks	Y + ks
10	0.563	5.41	0.035	0.037	.002	.072
20	0.344	5.10	0.033	0.089	.056	.122
30	0.224	4.89	0.032	0.141	.109	.173
50	0.284	5.00	0.033	0.245	.212	.278
75	0.918	5.74	0.037	0.375	.338	.412

t	$s(y_o) = (1 + \frac{1}{N'})s$	$ts(y_o)$	$Y - ts(y_o)$	$Y + ts(y_o)$
2.353	.0100	.024	.013	.061
2.353	.0087	.021	.068	.110
2.353	.0080	.019	.122	.160
2.353	.0083	.020	.225	.265
2.353	.0125	.029	.346	.404

N = 25 Observations, Five Calibrations

x	$\frac{1}{N'} = \frac{1}{25} + \frac{(x - 37)^2}{10,050}$	k	ks*	Y*	Y - ks	Y + ks
10	0.133	2.57	0.017	0.037	0.020	0.054
20	0.069	2.52	0.016	0.089	0.073	0.105
30	0.045	2.49	0.016	0.141	0.125	0.157
50	0.057	2.50	0.016	0.245	0.229	0.261
75	0.184	2.64	0.017	0.375	0.358	0.392

t	$s(y_o) = (1 + \frac{1}{N'})s$	$ts(y_o)$	$Y - ts(y_o)$	$Y + ts(y_o)$
1.71	.0072	.012	.025	.049
1.71	.0069	.012	.077	.101
1.71	.0068	.012	.129	.153
1.71	.0069	.012	.233	.257
1.71	.0077	.013	.362	.388

---

*The same values of Y and s were used in both computations since there were no actual data for five calibrations.

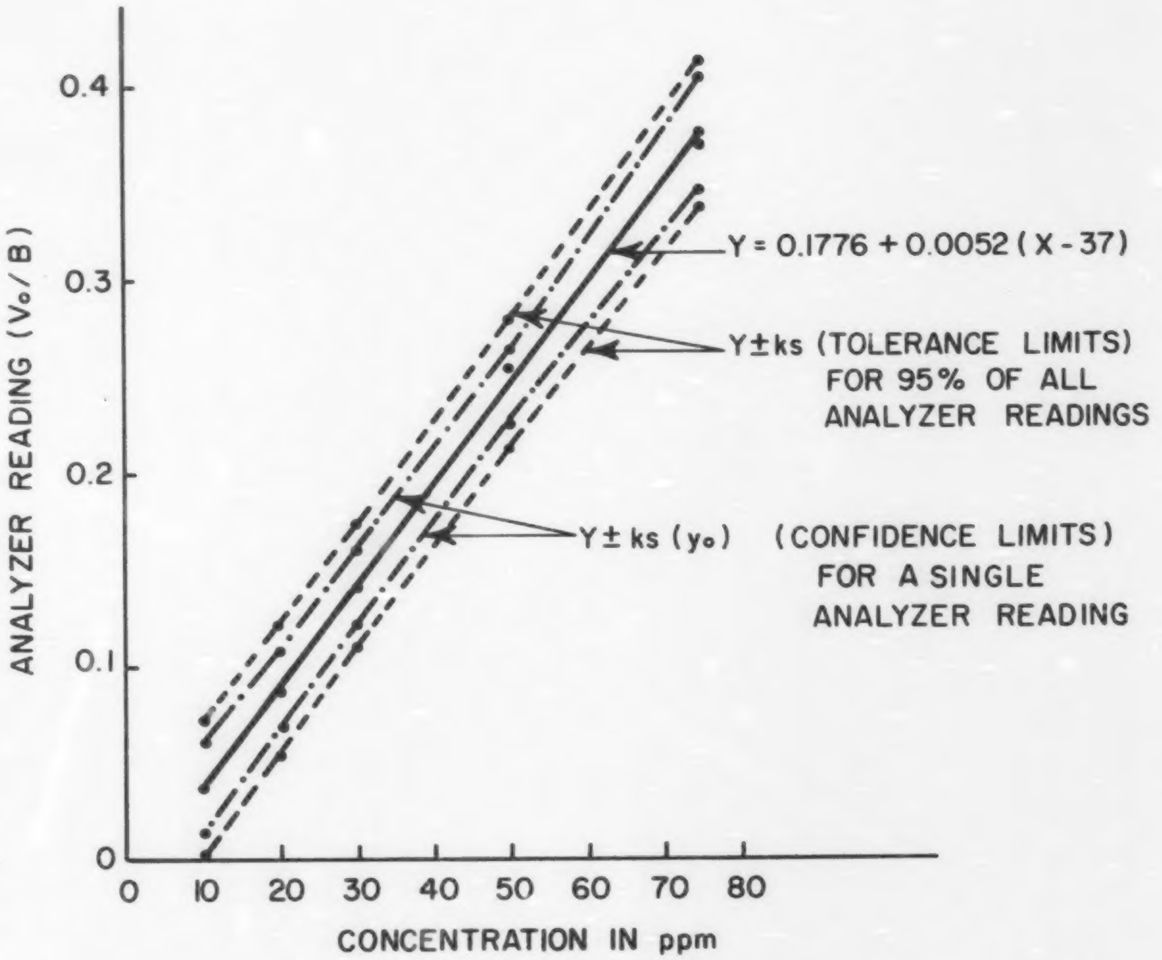


Figure 4-2: Confidence and Tolerance Limits for the Example.

MEAN AND VARIANCE OF NONLINEAR FUNCTIONS

In the previous examples, the variable of interest was expressed as a linear function of several variables with known coefficients. However, it is often necessary to estimate the mean and variance of a nonlinear combination such as that for the concentration of suspended particles in air as given below,

$$SP = \frac{(w_f - w_i) \times 10^6}{\frac{Q_i + Q_f}{2} \times T}$$

where SP denotes the concentration in $\mu g/m^3$,

w_f = final weight of filter in grams,

w_i = initial (tare) weight of filter in grams,

Q_i = initial flow rate in m^3/min,

Q_f = final flow rate in m^3/min, and

T = sampling time in min.

For relatively small variations (errors of measurement) in the variables w_f, w_i, ..., T in the above expression, it is usually sufficient to obtain an approximate linear expression of SP as a function of each of the variables and then proceed to use the formulas given previously. To obtain a linear approximation, differentiate SP with respect to each of the variables and expand the function in a Taylor Series as follows:

$$SP \approx SP\Big|_{mean} + \frac{\partial(SP)}{\partial\, w_f}\Big|_{mean} dw_f + \frac{\partial(SP)}{\partial w_i}\Big|_{mean} dw_i + \ldots + \frac{\partial(SP)}{\partial T}\Big|_{mean} dT$$

where $SP\Big|_{mean}$ denotes the value of SP where each variable is replaced by its mean or expected value,

$\frac{\partial(SP)}{\partial X}\Big|_{mean}$ denotes the value of the partial derivative of SP with respect to variable "X" and evaluated for each variable at its mean value, and

dx denotes the deviation of the variable "X" from its mean.

The variable X may be any one of w_f, w_i, ..., T. Using the results of equations (1) and (2) above, the approximate mean and variance of SP is obtained by

$$(5) \qquad \text{Mean (SP)} \simeq SP\Big|_{\text{mean}} + 0 \ ,$$

$$(6) \ \text{Variance (SP)} = \left(\frac{\partial(SP)}{\partial w_f}\Bigg|_{\text{mean}}\right)^2 \sigma_{w_f}^{\ 2} + \left(\frac{\partial(SP)}{\partial w_i}\Bigg|_{\text{mean}}\right)^2 \sigma^2_{w_i}$$

$$+ \ldots + \left(\frac{\partial(SP)}{\partial T}\Bigg|_{\text{mean}}\right)^2 \sigma^2_T$$

provided the deviations dw_f, dw_i, ..., dT are uncorrelated or independent. If they are not uncorrelated, one would use equation (3) and also have estimates of the correlations. An estimate of the variance (SP) is obtained by substituting estimates of the σ's throughout equation (6).

Another useful approach to obtaining the estimated variance of a function such as that for SP is to first take the logarithm of SP and then differentiate the resulting expression. In this case we obtain

$$\frac{d(SP)}{SP} = \frac{d(w_f - w_i)}{w_f - w_i} - \frac{d(Q_i + Q_f)}{Q_i + Q_f} - \frac{dT}{T} \ .$$

The variance of an expression

$$\frac{dX}{X}$$

is given by $\dfrac{\sigma^2_X}{\text{Mean}^2(X)} = CV^2(X)$, i.e., the square of the coefficient of variation. Hence, the variance of SP is given by

$$(7) \quad CV^2(SP) = CV^2(W_f - W_i) + CV^2(Q_i + Q_f) + CV^2(T)$$

where

$$(8) \quad \sigma^2(SP) = CV^2(SP)\ \text{Mean}^2(SP).$$

This approach to obtaining an estimate of the variance of SP is particularly useful if the CV's are expected to be constant (i.e., the standard deviation is some fixed percentage of the mean) and/or when the expression is a product (or quotient) of several variables.

The differentiation and algebraic calculations become very tedious for several variables and complex expressions. A computer program has been written not only to do the above computations using the first approach (the computer uses a numerically obtained derivative), but also to make certain checks on the adequacy of the approximation. Another computer program, in parallel with this one, simulates the measurement process by using assumed distributions of the variables in the expression and empirically obtains a distribution of SP. This approach was used in studying the distribution of SP as a function of several variables. The expression given above for SP may be expanded to include variables other than just those given. For example, the final weight of the filter might be expressed as a function of the pH of the filter surface and the weight of the collected particles. The concentration of the particles will vary with time of day, flow rate, etc. Hence, the expression can be made more specific and be useful provided the appropriate data are available to perform the analysis. These computer programs are fully described in Appendices 1 and 2.

An example computation is given below for illustrating the computational approach with equations (4), (5) and (6).

Parameter	Assumed Means	Assumed Standard Deviations
w_f	4031.2 mg	0.7 mg
w_i	3700 mg	0.7 mg
Q_i	1.20 m^3/min	0.03 m^3/min
Q_f	1.10 m^3/min	0.03 m^3/min
T	24 × 60 = 1440 min	7 min

$$SP = \frac{331.2 \times 10^3}{\frac{(1.2 + 1.1)}{2} \times 1440} = 200\ \mu g/m^3$$

Computation of Derivatives

$$\left.\frac{\partial(SP)}{\partial w_f}\right|_{mean} = \left.\frac{10^3}{\left(\frac{Q_i + Q_f}{2}\right) T}\right|_{mean} = .604 = - \left.\frac{\partial(SP)}{\partial w_i}\right|_{mean},$$

$$\left.\frac{\partial(SP)}{\partial Q_i}\right|_{mean} = \left.\frac{\partial(SP)}{\partial Q_f}\right|_{mean} = -\frac{2 \times 10^3 (w_f - w_i)}{T(Q_i + Q_f)^2} = -87.0$$

$$\frac{\partial(SP)}{\partial T} = -\frac{2 \times 10^3 (w_f - w_i)}{(Q_i + Q_f)T^2} = -.139 .$$

Thus,

$$SP \approx 200 + 0.604\ dw_f - 0.604\ dw_i - 87\ dQ_i - 87\ dQ_f$$

$$-0.139\ dT.$$

Hence,

$$\mu(SP) \approx 200\ \mu g/m^3,$$

$$\sigma^2(SP) \approx (0.604)^2\ (0.7)^2 + (0.604)^2\ (0.7)^2 + (87)^2\ (0.03)^2$$

$$+ (87)^2\ (0.03)^2 + (.139)^2\ 7^2,$$

$$\sigma(SP) = 3.86\ \mu g/m^3 .$$

Using formulas (7) and (8)

$$CV^2(SP) \approx (0.003)^2 + (0.0183)^2 + (0.005)^2$$

$$= 0.000368$$

$$CV(SP) = 0.0192$$

$$\sigma(SP) = CV(SP) \times \mu(SP) = (.0192)(200) = 3.84\ \mu g/m^3$$

which is the same as the above except for roundoff errors.

APPENDIX 5 CONTROL CHARTS FOR USE IN IMPROVING QUALITY OF MEASUREMENT PROCESS

There are several potential applications of quality control chart procedures which can be applied to aid in detecting deviations in the measurement process from expected values. For example, it may be desired to use a control chart for zero and span calibration checks. When dealing with measurements, it is customary to exercise control over the average quality of a process as well as its variability. Control over the average quality is accomplished by plotting the means of periodic samples on a so-called control chart for means, or an $\bar{x}$ chart. Variability is controlled by plotting the sample ranges or standard deviations on an R chart or a σ chart, respectively, depending on which statistic is used to estimate the population standard deviation. A control chart consists of a central line corresponding to the average quality at which the process is to perform, and lines corresponding to the upper and lower control limits. These limits are chosen so that values falling between them can be attributed to chance, while values falling outside them are interpreted as indicating a lack of control. Hence, if a zero or span check point deviates from the mean value by more than expected, action should be taken to determine and correct the cause.

The control limits can be established from prior knowledge of the process, level of error that is acceptable, or by collecting data over a period of time and letting the process establish its own limits as to what can be met in actual practice. Some useful quality control charts techniques are discussed briefly below.

STANDARD CONTROL CHART

No attempt will be made to give the detailed background concerning the charting techniques. This information is available in many textbooks (Refs. 22, 26, 29). Also see Refs. 25, 30 for discussion of specific charts and comparisons of their efficiencies. The final results are given here without background material.

Let the notation be as follows:

n = sample size,

σ = standard deviation of a single measurement (in the applications to pollutant measurements the standard deviation is usually the combination of among-day effects and within-day effects--e.g., see analyses of NO_2 and O_3 data in the respective sections of this report),

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ = standard deviation (error) of the average of n measurements,

μ = mean or expected value of the measurements (central line of the control chart),

k = multiple of $\sigma_{\bar{x}}$ used in obtaining the limits of the control chart for the average of n measurements,

R = range of the sample of n measurements, the largest value less the smallest value,

d = multiple of σ to be used in the construction of the control chart for ranges,

UCL = upper control line,

LCL = lower control line, and

$\bar{R}$ = mean value of the range R of k samples of size n drawn from a normal population.

Using the above definitions and information, control charts for both the mean and ranges can be readily obtained. These two charts are illustrated below in Figure 5-1 and 5-2. The appropriate factors for construction of the charts are given in Table 5-1 below.

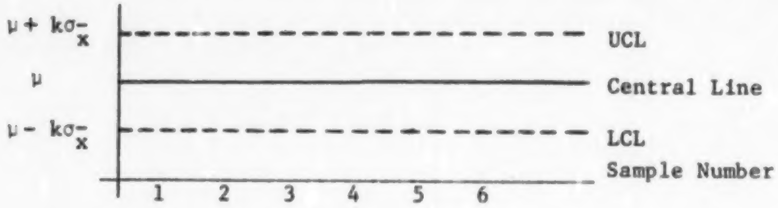


Figure 5-1: Control Chart for Means

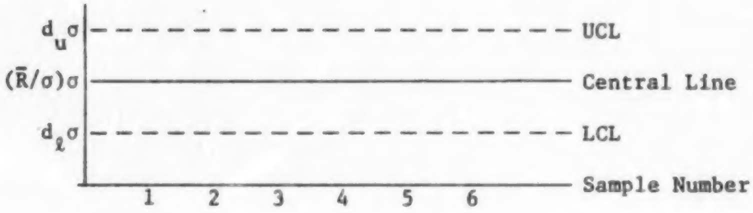


Figure 5-2: Control Chart for Ranges

To illustrate the construction of a control chart for ranges as shown in Figure 5-2, assume a sample size of 7 (i.e., n=7). From Table 5-1 the value of $(\bar{R}/\sigma)$ for n=7 is 2.70, and 2.70σ becomes the value of the central line of the control chart. Obtain values for d_u and d_ℓ for computing UCL and LCL, respectively, by specifying the upper and lower percentile to be used. If the 95th percentile is specified for the UCL, Table 5-1 shows that d_u has a value of 4.17 for a sample size of 7. The UCL is

$$d_u\sigma = 4.17\sigma.$$

Using 05 as the specified percentile for the LCL, Table 5-1 gives a value of 1.44 for d_ℓ, or 1.44σ for the LCL.

Table 5-1: TABLE OF FACTORS FOR CONSTRUCTING CONTROL CHARTS

		Values of d for Selected Percentiles			
		d_u		d_ℓ	
Sample size (n)	$\bar{R}/\sigma$	95	99	05	01
2	1.13	2.77	3.64	.09	.02
3	1.69	3.31	4.12	.43	.19
4	2.06	3.63	4.40	.76	.43
7	2.70	4.17	4.88	1.44	1.05

Values of k for Selected Percentiles			
	95	99	99.7
k(two-sided)	1.96 (2)	2.58	3.00
k(one-sided)	1.645	2.33	2.75

If for a given sample the mean $\bar{x}$ exceeds the upper control limit and/or if the range R exceeds the control limit $d_u\sigma$, the inference is made that the magnitude of the measurements and/or their dispersion, respectively, has increased. The risk of an incorrect inference that the measurement process is not in control when in fact it is stable or in control is controlled by the risk level selected according to Table 5-1.

There is also the risk of inferring that the process is still in control when in fact its level of dispersion has changed. This risk can be controlled by increasing the sample size or frequency of sampling and can be calculated, see e.g., Ref. 26.

RUN LENGTH AND RUN-SUM TESTS

There are frequent applications in the literature which suggest supplementing the standard quality control chart with either a run test or a run-sum test. The run test simply counts the number of consecutive values exceeding, e.g., the mean level. In the case of the chart for the mean, a run of length r would occur when there are r consecutive measurements exceeding the mean level, μ. If $r = 7$, then it is usually inferred that the mean has changed. Note that a run of seven, either below or above the mean, should occur with probability $2 \times (1/2)^7 = 1/64$ if there is, in fact, no change.

Another test which is suggested is a run-sum test. The standard quality control chart for the mean is subdivided into 8 zones, say; and an integral score is assigned to each zone weighted according the distance of the zone from the mean. See Figure 5-3 below illustrating the assignment of a score.

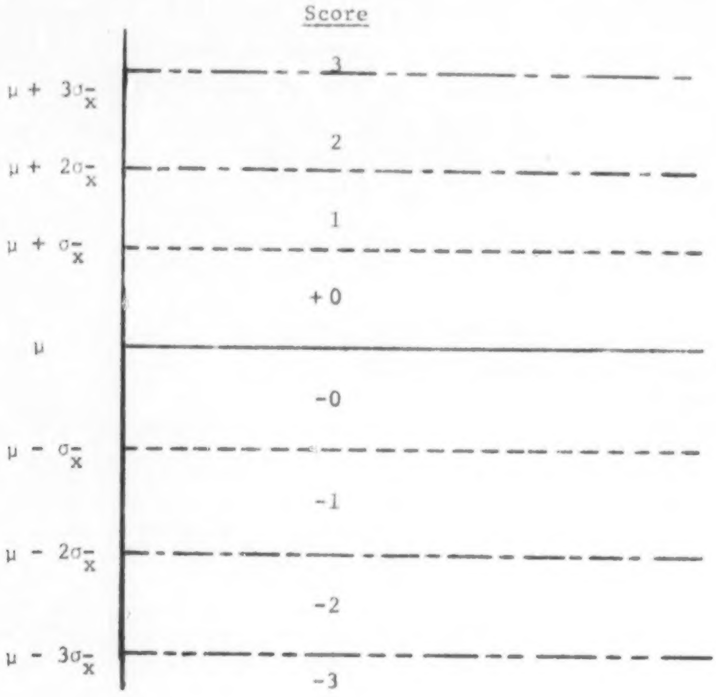
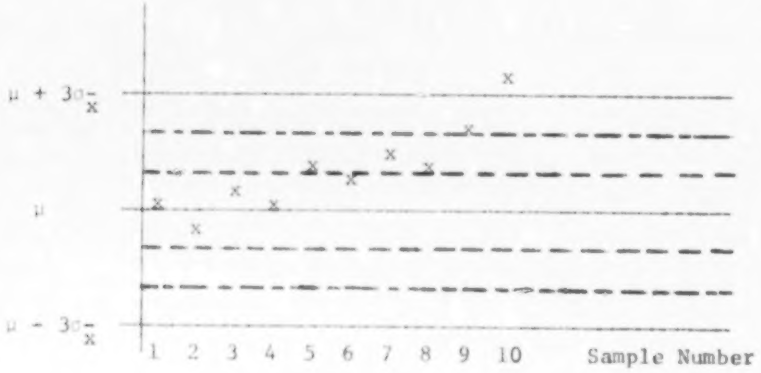

Score
3
$\mu + 3\sigma_{\bar{x}}$
2
$\mu + 2\sigma_{\bar{x}}$
1
$\mu + \sigma_{\bar{x}}$
+0
μ
-0
$\mu - \sigma_{\bar{x}}$
-1
$\mu - 2\sigma_{\bar{x}}$
-2
$\mu - 3\sigma_{\bar{x}}$
-3

Figure 5-3: Subdivision of Control Chart Into Zones and Assigned Scores.

If a run of positive scores occurs and adds to 5, say, then it is inferred that a positive shift in the mean measurement has occurred. Similarly for a negative run, hence the reason for the 0^+ and 0^- values. A run is interrupted once a sign change occurs, and the score returns to 0. The following figure illustrates the use of such a chart.



Run-Sum Score	+0	-0	+0	+0	1	1	2	3	5	8
Run-Length	1	1	1	2	3	4	5	6	7	8

Figure 5-4: Standard Control Chart Comparing Techniques

It is observed in Figure 5-4 that the run-length and run-sum tests indicate out of control at sample number 9, and the standard chart so indicates at sample number 10. These supplementary tests are easy to maintain and aid in detecting out of control when there are trends toward the limits without exceeding them. Reference 30 compares these techniques of quality control for abrupt changes in the process mean, μ. It is our desire to detect gradual changes which may result from the drift in the data points caused by instrument drift or component degradation. It is expected that the general comparisons given in Reference 30 would still be applicable.

One useful application of these quality control techniques would be in zero and span checks. For example, an operational rule might be to not make adjustments in the readings of concentrations of O_3 from the strip-chart unless the zero and/or span calibration points fall outside the range or limits given by a quality control chart for these readings. If the zero or span checkpoints do fall outside the limits, one would take action to determine if some degradation has occurred and/or make an appropriate adjustment in the calibration, e.g., complete the multipoint calibration and use the new calibration. This approach would avoid the result of "over correcting" the measurement process, the latter would yield resulting concentrations which would be more variable than the original uncorrected values. For example, the extreme case of making a daily adjustment in the mean based on measurement of a standard or control sample, where the standard deviation of the measurement of the standard is the same as that for an unknown sample, would result in a variance twice as large for the predicted values for the calculated concentrations of the unknown samples. To see that this is the case, consider the following simple example. Let the measured concentration of the unknown sample on the ith day be $y_i = \mu_u + \varepsilon_i$ and that for the standard sample

$$x_i = \mu_s + \delta_i .$$

Suppose the concentration of the unknown sample is reported as $y_i - x_i$; i.e., it is corrected for the standard sample. Now y_i has variance σ_ε^2, and x_i has variance σ_δ^2. Assuming these to be independent, we have

$$\sigma^2(y_i - x_i) = \sigma_\varepsilon^2 + \sigma_\delta^2 = 2\sigma^2, \text{ if } \sigma_\delta^2 = \sigma_\varepsilon^2 = \sigma^2 .$$

If y_i is corrected using an average of n, x_i values, e.g., for the last n days, then $y_i - \bar{x}$ has variance $\sigma^2 + \frac{\sigma^2}{n}$ under the assumption that $\sigma_\delta^2 = \sigma_\varepsilon^2$.

More complex considerations can be treated as suggested in Reference 25.